[Name of document] Specification

[Title of the Invention] A novel polypeptide, a method for producing it, cDNA encoding it, a vector carrying the cDNA, a host cell transformed with the vector, an antibody against the polypeptide and a pharmaceutical composition containing the polypeptide or the antibody

[Claims]

[Claim 1] A substantially purified form of the polypeptide comprising the amino acid sequence shown in SEQ ID NOS. 1, 4, 6, 9 or 12, homologue thereof, fragment thereof or homologue of the fragment.

[Claim 2] A polypeptide according to claim 1 comprising the amino acid sequence shown in SEQ ID NOS. 1, 4, 6, 9 or 12.

[Claim 3 A cDNA encoding the polypeptide according to claim 1.

[Claim 4] A cDNA according to claim 3 comprising the nucleotide sequence shown in SEQ ID NOS. 2, 5, 7, 10 or 13, or a fragment cDNA selectively hybridized to the cDNA.

[Claim 5] A cDNA according to claim 3 comprising the nucleotide sequence shown in SEQ ID NOS. 3, 8, 11 or 14, or a fragment cDNA selectively hybridized to the cDNA.

[Claim 6] A replication or expression vector carrying the cDNA according to claims 3 to 5.

[Claim 7] A host cell transformed with the replication or expression vector according to claim 6.

[Claim 8] A method for producing the polypeptide according to claim 1 or 2

which comprises culturing a host cell according to claim 7 under a condition effective to express the polypeptide according to claim 1 or 2.

[Claim 9] A monoclonal or polyclonal antibody against the polypeptide according to claim 1 or 2.

[Claim 10] A pharmaceutical composition containing the polypeptide according to claim 1 or 2 or the antibody according to claim 9, in association with pharmaceutically acceptable diluent and/or carrier.

[Detailed description of the Invention]

[Technical Field of the Invention]

The present invention relates to a novel polypeptide, a method for preparation of it, a cDNA encoding it, a vector containing the cDNA, a host cell transformed with the vector, an antibody against the peptide, and a pharmaceutical composition containing the polypeptide or the antibody.

[Problem to be dissolved by the Invention]

The present inventors et al. have diligently performed certain investigation in order to isolate novel factors (polypeptides) useful for treatment, diagnosis and/or study, particularly, secretory proteins containing secretory signal and membrane protein.

[Background of the Invention]

Until now, when a man skilled in the art intends to obtain a particular polypeptide or a cDNA encoding it, he generally utilizes methods by confirming an aimed biological activity in a tissue or in a cell medium, isolating and purifying the polypeptide and then cloning a gene or methods by "expression-cloning" with the guidance of the said biological activity. However, physiologically active polypeptides in living body have often many kinds of activities. Therefore, it happens increasingly that after cloning a

gene, the isolated gene is found to be identical to that encoding a polypeptide already known. In addition, some factors could be generated in only a very slight amount and/or under specific conditions and it makes difficult to isolate and to purify the factor and to confirm its biological activity.

[Related Arts]

Recent rapid developments in techniques for constructing cDNAs and sequencing techniques have made it possible to quickly sequence a large amount of cDNAs. By utilizing these techniques, a process, which comprises constructing cDNAs library using various cells or tissues, cloning the cDNA at random, identifying the nucleotide sequences thereof, expressing novel polypeptides encoded by them, is now in progress. Although this process is advantageous in that a gene can be cloned and information regarding its nucleotide sequence can be obtained without any biochemical or genetic analysis, the target gene can be discovered thereby only accidentally in many cases.

[Mean to dissolve the problem]

The present inventors have studied cloning method to isolate genes encoding proliferation and/or differentiation factors functioning in hematopoietic systems and immune systems. Focusing their attention on the fact that most of the secretory proteins such as proliferation and/or differentiation factors (for example various cytokines) and membrane proteins such as receptors thereof (hereafter these proteins will be referred to generally as secretory proteins and the like) have sequences called signal peptides in the N-termini, the inventors have conducted extensive studies on a process for efficiently and selectively cloning a gene encoding for a signal peptide. Finally, we have successfully developed a screening method for the signal peptides (signal sequence trap (SST)) by using mammalian cells (See Japanese Patent Application No. Hei 6-13951). We also developed yeast SST method on the same concept. By the method based on the same

conception using yeast, (yeast SST method), genes including sequence encoding signal peptide can be identified more easily and efficiently (See USP No. 5, 536, 637).

By using the present invention, the present inventors et al. achieved to find novel secretory proteins and membrane proteins produced from cell lines and tissue, for example, human adult brain tissue, cell lines derived from human brain tissue and cell line derived from human bone marrow, and cDNAs encoding them, and then completed the present invention.

The present invention provides the cDNA sequences identified as clones OC001, OM237, OA004b which were isolated by the said yeast SST method using cDNA libraries prepared from human adult brain tissue and cell lines derived from human brain tissue (T98G, ATCC No. CRL-1690). Clones OC001, OM237, OA004b were full-length cDNA including full cDNA sequences encoding membrain proteins (Each protein is represented as OC001, OM237, OA004b protein, respectively).

It was indicated from the results of homology search for the public database of the nucleic acid sequences by using BLASTN and FASTA, and for the public database of the amino acid sequences by using BLASTX, BLASTP and FASTA, that there was no sequence identical to the polypeptide sequence and the nucleotide sequences of OC001, OM237, OA004b of the present invention. In addition, the polypeptides of the present invention were expected to possess the transmembrane region by hydrophobisity analysis of the obtained amino acid sequences. From these results, it was proved that polypeptides OC001, OM237, OA004b of the present invention were new membrane proteins.

The present invention provides the cDNA sequence identified as clone OAF075b which was isolated by the said yeast SST method using cDNA libraries prepared from human bone marrow cell line HAS303 (human bone marrow cell line: provided from Prof. Keisuke Sotoyama, Dr. Makoto Aizawa, First Medicine, Tokyo Medical College. see J. Cell. Physiol. 148, 245-251, 1991 and Experimental Hematol. 22, 482-487, 1994). Clone

OAF075b was a full-length cDNA including a full cDNA sequence encoding secretory protein (this protein is represented as OAF075b protein).

It was indicated from the results of homology search for the public database of the nucleic acid sequences by using BLASTN and FASTA, and for the public database of the amino acid sequences by using BLASTX, BLASTP and FASTA, that there was no sequence identical to the polypeptide sequence and the nucleotide sequence of OAF075b of the present invention. In addition, the polypeptide of the present invention was expected to possess no transmembrane region by hydrophobisity analysis of the obtained amino acid sequence. From these results, it was proved that polypeptide of the present invention was a new secretory protein.

[The structure of the Invention]

The present invention relates to

(1) a polypeptide comprising an amino acid sequence of SEQ ID NOS. 1, 4, 6, 9 or 12,

(2) a cDNA encoding the polypeptide described in (1),

(3) a cDNA comprising a nucleotide sequence of SEQ ID NOS. 2, 5, 7, 10 or 13, and

(4) a cDNA comprising a nucleotide sequence of SEQ ID NOS. 3, 8, 11 or 14.

The present invention relates to a substantially purified form of the polypeptide comprising the amino acid sequence shown in SEQ ID NOS. 1, 4, 6, 9 or 12, homologue thereof, fragment thereof or homologue of the fragment.

Further, the present invention relates to cDNAs encoding the above peptides. More particularly the invention is provided cDNAs comprising nucleotide sequence shown in SEQ ID NOS. 2, 5, 7, 10 or 13, and cDNA containing a fragment which is selectively hybridizing to the cDNA comprising nucleotide sequence shown in SEQ ID NOS. 2, 5, 7, 10, 13, 3, 8, 11 or 14. A said cDNA capable for hybridizing to the cDNA includes the contemporary

sequence of the above sequence.

A polypeptide comprising amino acid sequence shown in SEQ ID NOS. 1, 4, 6, 9 or 12 in substantially purified form will generally comprise the polypeptide in a preparation in which more than 90%, e.g. 95%, 98% or 99% of the polypeptide in the preparation is that of the SEQ ID NOS. 1, 4, 6, 9 or 12.

A homologue of polypeptide comprising amino acid sequence shown in SEQ ID NOS. 1, 4, 6, 9 or 12 will be generally at least 70%, preferably at least 80 or 90% and more preferably at least 95% homologous to the polypeptide comprising the said amino acid sequence over a region of at least 20, preferably at least 30, for instance 40, 60 or 100 more contiguous amino acids. Such a polypeptide homologue will be referred to a polypeptide of the present invention.

Further, a fragment of polypeptide comprising amino acid sequence shown in SEQ ID NOS. 1, 4, 6, 9 or 12 or its homologues will be at least 10, preferably at least 15, for example 20, 25, 30, 40, 50 or 60 amino acids in length.

A cDNA capable of selectively hybridizing to the cDNA comprising nucleotide sequence shown in SEQ ID NOS. 2, 5, 7, 10, 13, 3, 8, 11 or 14 will be generally at least 70%, preferably at least 80 or 90% and more preferably at least 95% homologous to the cDNA comprising the said nucleotide sequence over a region of at least 20, preferably at least 30, for instance 40, 60 or 100 or more contiguous nucleotides. Such a cDNA will be referred to "a cDNA of the present invention".

Fragments of the cDNA comprising nucleotide sequence shown in SEQ ID NOS. 2, 5, 7, 10, 13, 3, 8, 11 or 14 will be at least 10, preferably at least 15, for example 20, 25, 30 or 40 nucleotides in length, and will be also referred to "a cDNA of the present invention" as used herein.

A further embodiment of the present invention provides replication and expression vectors carrying cDNA of the present invention. The vectors may be, for example, plasmid, virus or phage vectors provided with an origin

of replication, optionally a promoter for the expression of the said cDNA and optionally a regulator of the promoter. The vector may contain one or more selectable marker genes, for example ampicillin resistance gene. The vector may be used in vitro, for example of the production of RNA corresponding to the cDNA, or used to transfect a host cell.

A further embodiment of the present invention provides host cells transformed with the vectors for the replication and expression of the cDNA of the present invention, including the cDNA comprising nucleotide sequence shown in SEQ ID NOS. 2, 5, 7, 10, 13, 3, 8, 11 or 14 or the open reading frame thereof. The cells will be chosen to be compatible with the vector and may for example be bacterial, yeast, insect cells or mammalian cells.

A further embodiment of the present invention provides a method of producing a polypeptide which comprises culturing host cells of the present invention under conditions effective to express a polypeptide of the present invention. Preferably, in addition, such a method is carried out under conditions in which the polypeptide of the present invention is expressed and then produced from the host cells.

cDNA of the present invention may also be inserted into the vectors described above in an antisense orientation in order to prove for the production of antisense RNA. Such antisense RNA may be used in a method of controlling the levels of a polypeptide of the present invention in a cell.

The invention also provides monoclonal or polyclonal antibodies against a polypeptide of the present invention. The invention further provides a process for the production of monoclonal or polyclonal antibodies to the polypeptides of the present invention. Monoclonal antibodies may be prepared by common hybridoma technology using polypeptides of the present invention or fragments thereof, as an immunogen. Polyclonal antibodies may also be prepared by common means which comprise inoculating host animals, (for example a rat or a rabbit etc.), with polypeptides of the present invention and recovering immune serum.

The present invention also provides pharmaceutical compositions containing a polypeptide of the present invention, or an antibody thereof, in association with a pharmaceutically acceptable diluent and/or carrier.

The polypeptide of the present invention specified in (1) includes that which a part of their amino acid sequence is lacking (e.g., a polypeptide comprised of the only essential sequence for revealing a biological activity in an amino acid sequence shown in SEQ ID NO. 1), that which a part of their amino acid sequence is replaced by other amino acids (e. g., those replaced by an amino acid having a similar property) and that which other amino acids are added or inserted into a part of their amino acid sequence, as well as those comprising the amino acid sequence shown in SEQ ID NOS. 1, 4, 6, 9 or 12.

As known well, there are one to six kinds of codon as that encoding one amino acid (for example, one kind of codon for Methionine (Met), and six kinds of codon for Leucine (Leu) are known). Accordingly, the nucleotide sequence of cDNA can be changed in order to encode the polypeptide having the same amino acid sequence.

The cDNA of the present invention, specified in (2) includes a group of every nucleotide sequence encoding polypeptides (1) shown in SEQ ID NOS. 1, 4, 6, 9 or 12. There is a probability that yield of a polypeptide is improved by changing a nucleotide sequence.

The cDNA specified in (3) is the embodiment of the cDNA shown in (2), and indicate the sequence of natural form.

The cDNA shown in (4) indicates the sequence of the cDNA specified in (3) with natural non-translational region.

cDNA carrying nucleotide sequence shown in SEQ ID NOS. 3, 8, 11 or 14 is prepared by the following method:

Brief description of Yeast SST method (see USP No. 5, 536, 637) is as follows.

Yeast such as Saccharomyces cerevisiae should secrete invertase into the medium in order to take sucrose or raffinose as a source of energy or

carbon. (Invertase is an enzyme to cleave raffinose into sucrose and melibiose, sucrose into fructose and glucose). It is known that many known mammalian signal sequence make yeast secrete its invertase. From these knowledge, SST method was developed as a screening method to find novel signal peptide which make it possible can to secrete yeast invertase from mammalian cDNA library. SST method uses yeast growth on raffinose medium as a marker. Non-secretory type invertase gene SUC2 (GENBANK Accession No. V 01311) lacking initiation codon ATG was inserted to yeast expression vector to prepare yeast SST vector pSUC2. In this expression vector, ADH promoter, ADH terminator (both were derived from AAH5 plasmid (Gammerer, Methods in Enzymol. 101, 192-201, 1983)), 2μ ori (as a yeast replication origin), TRP1 (as a yeast selective marker), ColE1 ori (as a E. Coli replication origin) and ampicillin resistance gene (as a drug resistance marker) were inserted. Mammalian cDNA was inserted into the upstream of SUC2 gene to prepare yeast SST cDNA library. Yeast lacking secretory type invertase, was transformed with this library. If inserted mammalian cDNA encodes a signal peptide, yeast could survive in raffinose medium as a result of restoring secretion of invertase. Only to culture yeast colonies, prepare plasmids and determine the nucleotide sequence of the insert cDNAs, it is possible to identify novel signal peptide rapidly and easily.

Preparation of yeast SST cDNA library is as follows:

(1) mRNA is isolated from the targeted cells, double-strand synthesis is performed by using random primer with certain restriction enzyme (enzyme I) recognition site,

(2) obtained double-strand cDNA is ligated to adapter containing certain restriction endonuclease (enzyme II) recognition site, differ from enzyme I, digested with enzyme I and fractionated in a appropriate size,

(3) obtained cDNA fragment is inserted into yeast expression vector on the upstream region of invertase gene which signal peptide is deleted and the library was transformed.

Detailed description of each step is as follows:

(1) mRNA is isolated from mammalian organs and cell lines stimulate them with appropriate stimulator if necessary) by known methods (Molecular Cloning (Sambrook, J., Fritsch, E. F. and Maniatis, T., Cold Spring Harbor Laboratory Press, 1989) or Current Protocol in Molecular Biology (F. M. Ausubel et al, John Wiley & Sons, Inc) if not remark especially).

HAS303 (human bone marrow stroma cell line: provide from Professor Keisuke Sotoyama, Dr. Makoto Aizawa of First Medicine, Tokyo Medical College; see J. Cell. Physiol., 148, 245-251, 1991 and Experimental Hematol., 22, 482-487, 1994) or human glioblastoma cell line TG98G (ATCC No. CRL-1690) are chosen as a cell line. Human adult brain is chosen as a tissue source. Double-strand cDNA synthesis using random primer is performed by known methods.

Any sites may be used as restriction endonuclease recognition site I which is linked to adapter and restriction endonuclease recognition site II which is used in step (2), if both sites are different each other. Preferably, XhoI is used as enzyme I and EcoRI as enzyme II.

In step (2), cDNA is created blunt-ends with T4 DNA polymerase, ligated enzyme II adapter and digested with enzyme I. Fragment cDNA is analyzed with agarose-gel electrophoresis (AGE) and is selected cDNA fraction ranging in size from 300 to 800 bp. As mentioned above, any enzyme may be used as enzyme II if it is not same the enzyme I.

In step (3), cDNA fragment obtained in step (2) is inserted into yeast expression vector on the upstream region of invertase gene which signal peptide is deleted. E. Coli was transformed with the expression vector. Many vectors are known as yeast expression plasmid vector. For example, YEp24 is also functioned in E. Coli. Preferably pSUC2 as described above is used.

Many host E. Coli strains are known for transformation, preferably DH10B competent cell is used. Any known transformation method is available, preferably it is performed by electropolation method.

Transformant is cultured by conventional methods to obtain cDNA library for yeast SST method.

However not every all of the clones do not contain cDNA fragment. Further all of the gene fragments do not encode unknown signal peptides. It is therefore necessary to screen a gene fragment encoding for an unknown signal peptide from the library.

That is to say, screening of fragments containing a sequence encoding an appropriate signal peptide is performed by transformation of the cDNA library into Saccharomyces cerevisiae (e. g. YT455 strain) which lack invertase (it may be prepared by known methods). Transformation of yeast is performed by known methods, e. g. lithium acetate method. Transformant is cultured in a selective medium, then transferred to a medium containing raffinose as a carbon source. Survival colonies are selected and then prepared plasmid. Survival colonies on a raffinose-medium indicates that some signal peptide of secretory protein was inserted to this clone.

As for isolated positive clones, the nucleotide sequence is determined. As to a cDNA encodes unknown protein, full-length clone may be isolated by using cDNA fragment as a probe and then determined to obtain full-length nucleotide sequence. These manipulation is performed by known methods.

Once the nucleotide sequences shown in SEQ ID NO. 2, 5, 7, 10 or 13 are determined partially or preferably fully, it is possible to obtain DNA encode mammalian protein itself, homologue or subset. cDNA library or mRNA derived from mammals was screened by PCR with any synthesized oligonucleotide primers or by hybridization with any fragment as a probe. It is possible to obtain DNA encodes other mammalian homologue protein from other mammalian cDNA or genome library.

If a cDNA obtained above contains a nucleotide sequence of cDNA fragment obtained by SST (or consensus sequence thereof), it will be thought that the cDNA encodes signal peptide. So it is clear that the cDNA will be full-length or almost full. (All signal peptides exist at N-termini of a protein

and are encoded at 5'-temini of open reading frame of cDNA)

The confirmation may be carried out by Northern analysis with the said cDNA as a probe. It is thought that the cDNA is almost complete length, if length of the cDNA is almost the same length of the mRNA obtained in the hybridizing band.

Once the nucleotide sequences shown in SEQ ID NOS. 2, 5, 7, 10 or 13 are determined, DNAs of the invention are obtained by chemical synthesis, or by hybridization making use of nucleotide fragments which are chemically synthesized as a probe. Furthermore, DNAs of the invention are obtained in desired amount by transforming a vector that contains the DNA into a proper host, and culturing the transformant.

The polypeptides of the present invention may be prepared by:

(1) isolating and purifying from an organism or a cultured cell,

(2) chemically synthesizing, or

(3) using recombinant cDNA technology,

preferably, by the method described in (3) in an industrial production.

Examples of expression system (host-vector system) for producing a polypeptide by using recombinant cDNA technology are the expression systems of bacteria, yeast, insect cells and mammalian cells.

In the expression of the polypeptide, for example, in E. Coli, the expression vector is prepared by adding the initiation codon (ATG) to 5' end of a cDNA encoding mature peptide, connecting the cDNA thus obtained to the downstream of a proper promoter (e. g., trp promoter, lac promoter, λPL promoter, T7 promoter etc.), and then inserting it into a vector (e. g., pBR322, pUC18, pUC19 etc.) which functions in an E. Coli strain.

Then, an E. Coli strain (e. g., E. Coli DH1 strain, E. Coli JM109 strain, E. Coli HB101 strain, etc.) which is transformed with the expression vector described above may be cultured in a appropriate medium to obtain the desired polypeptide. When a signal sequence of bacteria (e. g., signal sequence of pel B) is utilized, the desired polypeptide may be also released in periplasm. Furthermore, a fusion protein with other polypeptide may be

also produced readily.

In the expression of the polypeptide, for example, in a mammalian cells, for example, the expression vector is prepared by inserting the cDNA encoding nucleotide shown in SEQ ID NOS. 3, 8, 11 or 14 into the downstream of a proper promoter (e. g., SV40 promoter, LTR promoter, metallothionein promoter etc.) in a proper vector (e. g., retrovirus vector, papilloma virus vector, vaccinia virus vector, SV40 vector, etc.). A proper mammalian cell (e. g., monkey COS-7 cell, Chinese hamster CHO cell, mouse L cell etc.) is transformed with the expression vector thus obtained, and then the transformant is cultured in a proper medium to express the aimed secretory protein and membrane protein of the present invention by the following method.

In case of secretory protein as for the present invention, the aimed polypeptide was expressed in the supernatant of the cells. In addition, fusion protein may be prepared by conjugating cDNA fragment encoding the other polypeptide, for example, Fc portion of antibody.

On the other hand, in case of membrane protein as for the present invention, the aimed polypeptide was expressed on the cell membrane. A cDNA encoding the nucleotide sequence of SEQ ID NOS. 2, 5, 7, 10 or 13 with deletion of extracellular region was inserted into the said vector, transfected into the an adequate mammalian cells to secret the aimed soluble polypeptide in the culture medium. In addition, fusion protein may be prepared by conjugating cDNA fragment encoding the said mutant with deletion of extracellular region and other polypeptide, for example, Fc portion of antibody.

The polypeptide available by the way described above can be isolated and purified by conventional biochemical method.

[Effect of the Invention]

It is considered that the polypeptide of the present invention and a cDNA which encodes the polypeptide will show one or more of the effects or

biological activities (including those which relates to the assays cited below) The effects or biological activities described in relation to the polypeptide of the present invention are provided by administration or use of the polypeptide or by administration or use of a cDNA molecule which encodes the polypeptide (e. g., vector suitable for gene therapy or cDNA introduction).

[Cytokine activity and cell proliferation/differentiation activity]

The protein of the present invention may exhibit cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of a polypeptide of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines.

[Immune stimulating/suppressing activity]

The protein of the present invention may also exhibit immune stimulating or immune suppressing activity. The protein of the present invention may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), e. g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral infection such as HIV as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases causes by viral, bacterial, fungal or other infection may be treatable using the polypeptide of the present invention, including infections by HIV, hepatitis viruses, herpes viruses, mycobacteria, leshmania, malaria and various fungal infections such as candida. Of course, in this regard, the protein of the

present invention may also be useful where a boost to the immune system generally would be indicated, i. e., in the treatment of cancer.

The protein of the present invention may be useful in the treatment of allergic reactions and conditions, such as asthma or other respiratory problems. The protein of the present invention may also be useful in the treatment of the other conditions required to suppress the immuno system (for example, asthma or respiratory disease)

The protein of the present invention may also suppress chronic or acute inflammation, such as, for example, that associated with infection such as septic shock or systemic inflammatory response syndrome (SIRS), inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-I wherein the effect was demonstrated by IL- 11.

[Hematopoiesis regulating activity]

The protein of the present invention may be useful in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell deficiencies. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis. The said biological activities are concerned with the following all or some example(s). e. g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemia or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i. e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in

place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either in-vitro or ex-vivo (i. e. in conjunction with bone marrow transplantation) as normal cells or genetically manipulated for gene therapy.

The activity of the protein of the present invention may, among other means, be measured by the following methods :

[Tissue generation/regeneration activity]

The protein of the present invention also may have utility in compositions used for bone, cartilage, tendon, Ligament and/or nerve tissue growth or regeneration, as well as for wound healing and tissue repair, and in the treatment of bums, incisions and ulcers.

The protein of the present invention, which induces cartilage and/or bone growth in circumstances where bone is not normally formed, may be applied to the healing of bone fractures and cartilage damage or defects in humans and other animals. Such a preparation employing the protein of the present invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

The protein of the present invention may also be used in the treatment of periodontal disease, and in other tooth repair processes. Such agents may provide an environment to attract bone-forming cells, stimulate growth of bone-forming cells or induce differentiation of progenitors of bone-forming cells. The protein of the present invention may also be useful in the

treatment of osteoporosis or osteoarthritis, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes.

Another category of tissue regeneration activity that may be attributable to the protein of the present invention is tendon/ligament formation. The protein of the present invention, which induces tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, may be applied to the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing the protein inducing a tendon/Ligament-like tissue may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide an environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon Ligament cells or progenitors ex vivo for return in vivo to effect tissue repair. The compositions of the present invention may also be useful in the treatment of tendinitis, Carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The protein of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue. i. e. for the treatment of central and peripheral nervous system diseases and neuropathies. as well as mechanical and traumatic disorders, which involve

degeneration, death or trauma to neural cells or nerve tissue. More specifically, the protein of the present invention may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using the polypeptide of the present invention.

It is expected that the protein of the present invention may also exhibit activity for generation of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the proliferation of cells comprising such tissues. Part of the desired effects may be by inhibition of fibrotic scarring to allow normal tissue to regenerate.

The protein of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

[Activin/Inhibin activity]

The protein of the present invention may also exhibit activin- or inhibin-related activities. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, the protein of the present invention alone or in heterodimers with a member of the inhibin *a family, may be useful as a

contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the protein of the present invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin-*b group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary (See USP 4, 798, 885). The protein of the present invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as cows, sheep and pigs.

[Chemotactic/chemokinetic activity]

The protein of the present invention may have chemotactic or chemokinetic activity e. g., functioning as a chemokine, for mammalian cells, including, for example, monocytes, neutrophils, T-cells, mast cells, eosinophils and/or endothelial cells. Chemotactic and chemokinetic proteins can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic proteins provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

If a protein or peptide can stimulate, directly or indirectly, the directed orientation or movement of such cell population, it has chemotactic activity for a particular cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

[Hemostatic and thrombolytic activity]

The protein of the present invention may also exhibit hemostatic or thrombolyic activity. As a result, such a protein is expected to be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A protein of the present invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom such as, for example, infarction or stroke.

[Receptor/ligand activity]

The protein of the present invention may also demonstrate activity as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including cellular adhesion molecules such as Selectins, Integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses. Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. The protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

[Other activity]

The protein of the present invention may also exhibit one or more of the following additional activities or effects: inhibiting growth of or killing the infecting agents including bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) body characteristics including height, weight, hair color, eye color, skin, other tissue pigmentation, or organ or body

part size or shape such as, for example, breast augmentation or diminution etc.; effecting elimination of dietary fat, protein, carbohydrate; effecting behavioral characteristics including appetite, libido, stress, cognition (including cognitive disorders), depression and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases.

The protein with above activities, is suspected to have following functions by itself or interaction with its ligands or receptors or association with other molecules. For example, proliferation or cell death of B cells, T cells and/or mast cells; specific induction by promotion of class switch of immunoglobulin genes; differentiation of B cells to antibody-forming cells; proliferation, differentiation, or cell death of precursors of granulocytes; proliferation, differentiation, or cell death of precursors of monocytes-macrophages; proliferation, of up regulation or cell death of neutrophils, monocytes-macrophages, eosinophils and/or basophils; proliferation, or cell death of precursors of megakaryocytes; proliferation, differentiation, or cell death of precursors of neutrophils; proliferation, differentiation, or cell death of precursors of T cells and B cells; promotion of production of erythrocytes; sustainment of proliferation of erythrocytes, neutrophils, eosinophils, basophils, monocytes-macrophages, mast cells, precursors of megakaryocyte; promotion of migration of neutrophils, monocytes-macrophages, B cells and/or T cells; proliferation or cell death of thymocytes; suppression of differentiation of adipocytes; proliferation or cell death of natural killer cells; proliferation or cell death of hematopoietic stem cells; suppression of proliferation of stem cells and each hematopoietic precursor cells; promotion of differentiation from mesenchymal stem cells to osteoblasts or chondrocytes, proliferation or cell death of mesenchymal stem cells, osteoblasts or chondrocytes and promotion of bone absorption by activation of osteoclasts and promotion of differentiation from monocytes to osteoclasts.

The present polypeptide is also suspected to function to nervous system, so expected to have functions below; differentiation to kinds of neurotransmitter-responsive neurons, survival or cell death of these cells; promotion of proliferation or cell death of glial cells; spread of neural dendrites; survival or cell death of gangriocytes; proliferation, promotion of differentiation, or cell death of astrocytes; proliferation, survival or cell death of peripheral neurons; proliferation or cell death of Schwann cells; proliferation, survival or cell death of motoneurons.

Furthermore, in the process of development of early embryonic, the present polypeptide is expected to promote or inhibit the organogenesis of epidermis, brain, backbone, and nervous system by induction of ectoderm, that of notochord connective tissues (bone, muscle, tendon), hemocytes, heart, kidney, and genital organs by induction of mesoderm, and that of digestive apparatus (stomach, intestine, liver, pancreas), respiratory apparatus (lung, trachea) by induction of endoderm. In adult, also, this polypeptide is thought to proliferate or inhibit the above organs.

Therefore, the polypeptide of the present invention itself is expected to be used as an agent for the prevention or treatment of disease of progression or suppression of immune, nervous, or bone metabolic function, hypoplasia or overgrowth of hematopoietic cells: for example, inflammatory disease (rheumatism, ulcerative colitis, etc.), decrease of hematopoietic stem cells after bone marrow transplantation, decrease of leukocytes, platelets, B-cells, or T-cells after radiation exposure or chemotherapeutic dosage against cancer or leukemia, anemia, infectious disease, cancer, leukemia, AIDS, bone metabolic disease (osteoporosis etc.), various degenerative disease (Alzheimer's disease, multiple sclerosis, etc.), or nervous lesion.

In addition, since the present polypeptide is thought to induce the differentiation or growth of organs derived from ectoderm, mesoderm, and endoderm, this polypeptide is expected to be an agent for tissue repair (epidermis, bone, muscle, tendon, heart, kidney, stomach, intestine, liver, pancreas, lung, and trachea, etc.).

By using polyclonal or monoclonal antibodies against the said polypeptide, quantitation of the said polypeptide in the body can be performed. It can be used in the study of relationship between this polypeptide and disease or diagnosis of disease, and so on. Polyclonal and monoclonal antibodies can be prepared using this polypeptide or its fragment as an antigen by conventional methods.

Identification, purification or molecular cloning of known or unknown proteins which bind the present polypeptide (preferably polypeptide of extracellular domain) can be performed using the said polypeptide by, for example, preparation of the affinity-column.

Identification of the downstream signal transmission molecules which interact with the said polypeptide in cytoplasma and molecular cloning of the gene can be performed by west-western method using the said polypeptide (preferably polypeptide of transmembrane region or intracellular domain), or by yeast two-hybrid system using the cDNA (preferably cDNA encoding transmembrane region or cytoplasmic domain of the polypeptide).

Agonists/antagonists of this receptor polypeptide and inhibitors between receptor and signal transduction molecules can be screened using the present polypeptide.

cDNAs of the present invention are useful not only the important and essential template for the production of the polypeptide of the present invention which is expected to be largely useful, but also be useful for diagnosis or therapy (for example, treatment of gene lacking, treatment to stop the expression of the polypeptide by antisense cDNA (mRNA)). Genomic cDNA may be isolated with the cDNA of the present invention, as a probe. As the same manner, a human gene encoding which can be highly homologous to the cDNA of the present invention, that is, which encodes a polypeptide highly homologous to the polypeptide of the present invention and a gene of animals excluding mouse which can be highly homologous to the cDNA of the present invention, also may be isolated.

[Application to Medicaments]

The polypeptide of the present invention or the antibody specific for the polypeptide of the present invention is administered systemically or topically and in general orally or parenterally, preferably parenterally, intravenously and intraventricularly, for preventing or treating the said diseases.

The doses to be administered depend upon age, body weight, symptom, desired therapeutic effect, route of administration, and duration of the treatment etc. In human adults, one dose per person is generally between 100 μg and 100 mg, by oral administration, up to several times per day, and between 10 μg and 100 mg, by parental administration up to several times per day.

As mentioned above, the doses to be used depend upon various conditions. Therefore, there are cases in which doses lower than or greater than the ranges specified above may be used.

The compounds of the present invention, may be administered as solid compositions, liquid compositions or other compositions for oral administration, as injections, liniments or suppositories etc. for parental administration.

Solid compositions for oral administration include compressed tablets, pills, capsules, dispersible powders, and granules. Capsules include soft or hard capsules.

In such compositions, one or more of the active compound(s) is or are admixed with at least one inert diluent (such as lactose, mannitol, glucose, hydroxypropyl cellulose, microcrystalline cellulose, starch, polyvinylpyrrolidone, magnesium metasilicate aluminate, etc.). The compositions may also comprise, as is normal practice, additional substances other than inert diluents: e. g. lubricating agents (such as magnesium stearate etc.), disintegrating agents (such as cellulose calcium glycolate, etc.), stabilizing agents (such as human serum albumin, lactose etc.), and assisting agents for dissolving (such as arginine, asparaginic acid etc.).

The tablets or pills may, if desired, be coated with a film of gastric or enteric materials (such as sugar, gelatin, hydroxypropyl cellulose or hydroxypropylmethyl cellulose phthalate, etc.), or be coated with more than two films. And then, coating may include containment within capsules of absorbable materials such as gelatin.

Liquid compositions for oral administration include pharmaceutically-acceptable emulsions, solutions, syrups and elixirs. In such compositions, one or more of the active compound(s) is or are contained in inert diluent(s) commonly used (purified water, ethanol etc.). Besides inert diluents, such compositions may also comprise adjuvants (such as wetting agents, suspending agents, etc.), sweetening agents, flavoring agents, perfuming agents, and preserving agents.

Other compositions for oral administration include spray compositions which may be prepared by known methods and which comprise one or more of the active compound(s). Spray compositions may comprise additional substances other than inert diluents: e. g. stabilizing agents (sodium sulfite etc.), isotonic buffer (sodium chloride, sodium citrate, citric acid, etc.). For preparation of such spray compositions, for example, the method described in the United States Patent No. 2, 868, 691 or 3, 095, 355 (herein incorporated in their entireties by reference) may be used.

Injections for parental administration include sterile aqueous or non-aqueous solutions, suspensions and emulsions. In such compositions, one or more active compound(s) is or are admixed with at least one inert aqueous diluent(s) (distilled water for injection, physiological salt solution, etc.) or inert non-aqueous diluents(s) (propylene glycol, polyethylene glycol, olive oil, ethanol, POLYSOLBATE 80 (Trade mark) etc.).

Injections may comprise additional compound other than inert diluents: e. g. preserving agents, wetting agents, emulsifying agents, dispersing agents, stabilizing agent (such as human serum albumin, lactose, etc.), and assisting agents such as assisting agents for dissolving (arginine, asparaginic acid, etc.).

[Examples]

The invention is illustrated by the following examples relating to clone OC001 of the present invention, but not limit the invention.

Preparation of Poly(A)$^+$RNA

Total RNA was prepared from human placenta tissue by TRIzol reagent (Trade Mark, marketed from GIBCOBRL Co.). Poly(A)$^+$RNA was purified from the total RNA by mRNA Purification Kit (Trade name, marketed from Pharmacia Co.).

Preparation of yeast SST cDNA library

Double strand cDNA was synthesized by Super Script Plasmid System for cDNA Synthesis and Plasmid Cloning (Trade name, marketed from GIBCOBRL Co.) with above poly(A)$^+$RNA as template and random 9mer as primer which was containing XhoI site:

5'-CGATTGAATTCTAGACCTGCCTCGAGNNNNNNNNN-3'

cDNA was ligated EcoRI adapter by DNA ligation kit ver. 2 (Trade name, marketed from Takara-Shuzo Co., this kit was used in all ligating steps hereafter) and digested by XhoI. cDNAs were separated by agarose-gel electrophoresis. 300～800 bp cDNAs were isolated and were ligated to EcoRI/NotI site of pSUC2 (see US Patent No. 5, 536, 637). E. Coli DH10B strains were transformed by pSUC2 with electropolation to obtain yeast SST cDNA library.

Screening by SST method and determination of nucleotide sequence of SST positive clone

Plasmids of the said cDNA library were prepared. Yeast YTK12 strains were transformed by the plasmids with lithium acetate method (Current Protocols In Molecular Biology 13.7.1). The transformed yeast were plated on triptphan-free medium (CMD-Trp medium) for selection. The

plate was incubated for 48 hour at 30 ℃. Replica of the colony (transformant) which was obtained by Accutran Replica Plater (Trade name, marketed from Schleicher & Schuell Co.) were placed onto YPR plate containing raffinose for carbon source, and the plate was incubated for 14 days at 30℃. After 3 days, each colony appeared was streaked on YPR plate again. The plates were incubated for 48 hours at 30℃. Single colony was inoculated to YPD medium and was incubated for 48 hours at 30℃. Then plasmids were prepared. Insert cDNA was amplified by PCR with two kind primers which exist end side of cloning site on pSUC2 (sense strand primers were biotinylated). Biotinylated single strand of cDNAs were purified with Dynabeads (Trade name, marketed from DYNAL Co.) and the nucleotide sequences were determined. Sequencing was performed by Dye Terminator Cycle Sequencing Ready Reaction with DNA Sequencing kit (Trade name, marketed from Applied Biosystems Inc.) and sequence was determined by DNA sequencer 373 (Applied Biosystems Inc.) (All sequencing hereafter was carried out with this method).

We tried to carry out cloning of full-length cDNA which was proved to be new one according to the homology search for the obtained nucleotide sequences and deduced amino acid sequences in data base.

Cloning of a full-length cDNA and determination of nucleotide

A full-length cDNA was cloned using Marathon cDNA Amplification Kit (Trade name, marketed from Clontech Co.) according to 3' RACE (Rapid Amplification of cDNA End) method. I. e., poly $(A)^+$RNA in human adult brain tissue 27mer primer OC001-F1:

5'-GTCCTTCAGCAAAACAGTGGATTTAAA-3'

containing the deduced initiation ATG codon region based on the information of nucleotide sequence obtained by SST, was prepared. PCR was performed with the said primer and adapter primer attached in the kit. A cDNA which was amplified with clone OC001 specifically, was separated with agarose-gel electrophoresis, ligated to pT7 Blue-2 T-Vector (Trade name,

marketed from Novagen Co) and transfected into E. Coli DH5α to prepare the plasmid. Nucleotide sequences of 5'-end were determined, and the existence of nucleotide sequence OC001 SST cDNA was confirmed. Nucleotide sequence of full-length OC001 SST cDNA was determined and then sequence shown in SEQ ID NO. 3 was obtained. An open reading frame was determined and deduced amino acid sequence and nucleotide sequence shown in SEQ ID NOS. 1, 2, 4 and 5, respectively, were obtained.

It was indicated from the results of homology search for the public database of the nucleic acid sequences by using BLASTN and FASTA, and for the public database of the amino acid sequences by using BLASTX, BLASTP and FASTA, that there was no sequence identical to the polypeptide sequence and the nucleotide sequence of OC001 of the present invention. In addition, the polypeptide of the present invention was expected to possess the transmembrane region at C-terminal and to be GPI anchor type by hydrophobisity analysis of the obtained amino acid sequence. From these results, it was proved that polypeptide of the present invention was new membrane protein. Further, the search using BLASTX, BLASTP and FASTA revealed a significant homology between clone OC001 (region of 12th～344th amino acid in SEQ ID NO. 1) and neurotrimin [Rattus norvegicus] (region of 9th～344th amino acid of Genbank Accession U16845) and opioid-binding cell adhesion molecule [Homo sapiens] (region of 9th～345th amino acid of Genbank Accession L34774). Based on these homologies, clone OC001 and nervous cell adhesion molecule family including neurotrimin and opioid-binding cell adhesion molecule were expected to share at least some activity.

In Example relating to clone OM237 of the present invention, the same procedure as in Example of OC001 was used except for the following points.
Cloning of a full-length cDNA and determination of nucleotide

A full-length cDNA was cloned by the same procedure as OC001 using Marathon cDNA Amplification Kit (Trade name, marketed from Clontech

Co.) according to 3'RACE. A double-strand cDNA was prepaed from RNA derived from each clone, i. e., poly(A)$^+$RNA of human adult brain tissue. 27mer primer OM237-F1:

5'-CCAGAAAGCACAGCCCTGATTCTGCGT-3'

containing the deduced initiation ATG codon region based on the information of nucleotide sequence obtained by SST, was prepared. PCR was performed with the said primer and adapter primer attached in the kit. A cDNA which was amplified with clone OM237 specifically, was recloned by the same method as OC001 to determine full necleotide sequence and obtain the sequence shown in SEQ ID NO. 8. An open reading frame was determined and deduced amino acid sequence and nucleotide sequence shown in SEQ ID NOS. 6 and 7, respectively, were obtained.

It was indicated from the results of homology search for the public database of the nucleic acid sequences by using BLASTN and FASTA, and for the public database of the amino acid sequences by using BLASTX, BLASTP and FASTA, that there was no sequence identical to the polypeptide sequence and the nucleotide sequence of OM237 of the present invention. In addition, the polypeptide of the present invention was expected to possess the transmembrane region by hydrophobisity analysis of the obtained amino acid sequence. From these results, it was proved that polypeptide of the present invention was new membrane protein.

In Example relating to clone OA004b of the present invention, the same procedure as in Example of OC001 was used except for the following points.

prepparation of poly(A)$^+$RNA

Total RNA was prepared from human glioblastoma cell line T98G (ATCC No. CRL-1690) by TRIzol reagent (Trade Mark, marketed from GIBCOBRL Co.). Poly(A)$^+$RNA was purified from total RNA by mRNA Purification Kit (Trade name, marketed from Pharmacia Co.).

Cloning of a full-length cDNA and detemination of amino acid sequnce

A full-length cDNA was cloned by GENETRAPPER cDNA Positive Selection System (GIBCOBRL Co.). First, dT-primed cDNA library was prepared using plasmid pSPORT1 (GIBCOBRL Co.) as a vector from poly(A)$^+$RNA of human glioblastoma cell line T98G by Super Script Plasmid System for cDNA Synthesis and Plasmid Cloning (Trade name, marketed from GIBCOBRL Co.). After preparing 27mer biotinylated primer OA004-F1:

5'-biotin-ATGCACATCTTCAAGCATGCTCAG-3',

based on the information of nucleotide sequence obtained by SST, plasmid hybridized specifically with the biotinylated primer were recovered from the cDNA library according to the method of Gene Trapper Kit and then transfected into E. Coli DH10B. Colony hybridization with OA004 SST cDNA which was labeled with ^{32}P-dCTP, as a probe, was performed by using Random Primer DNA Labeling kit (Trade name, marketed from Takara-Shuzo Co.) according to known method to isolate the positive clone and to prepare the plasmid. Full Nucleotide sequences was determined, and then sequence shown in SEQ ID NO. 11, which was named as OA004b, was obtained. An open reading frame was determined and deduced amino acid sequence and nucleotide sequence shown in SEQ ID NOS. 9 and 10, respectively, were obtained.

It was indicated from the results of homology search for the public database of the nucleic acid sequences by using BLASTN and FASTA, and for the public database of the amino acid sequences by using BLASTX, BLASTP and FASTA, that there was no sequence identical to the polypeptide sequence and the nucleotide sequence of OA004b of the present invention. In addition, the polypeptide of the present invention was expected to possess the transmembrane region by hydrophobisity analysis of the obtained amino acid sequence. From these results, it was proved that polypeptide of the present invention was new membrane protein. However, the search using BLASTX, BLASTP and FASTA revealed a significant homology between

clone OA004b (region of 171st～311st amino acid in SEQ ID NO. 9) and Hypothetical 52.8kD protein [Caenorhabdtis elegans] (region of 299th～453rd amino acid of Swiss Prot Accession YJ95_CAEEL), and between OA004b (region of 194th～277th amino acid in SEQ ID NO. 9) and presenilin-2 [Homo sapiens] (region of 340th～416th amino acid of Genbank Accession A56993). Based on these homologies, clone OA004b and presenilin family were expected to share at least some activity.

In Example relating to clone OAF075b of the present invention, the same procedure as in Example of OC001 was used except for the following points.

Preparation of poly(A)$^+$RNA

Total RNA was prepared from human bone marrow stroma cell line HAS303 (provided from Prof. Keisuke Sotoyama, Dr. Makoto Aizawa, First Medicine, Tokyo Medical College) by TRIzol reagent (Trade Mark, marketed from GIBCOBRL Co.). Poly(A)$^+$RNA was purified from the total RNA by mRNA Purification Kit (Trade name, marketed from Pharmacia Co.).

Cloning of a full-length cDNA and detemination of amino acid sequnce

A full-length cDNA was cloned by the same procedure as OC001 using Marathon cDNA Amplification Kit (Trade name, marketed from Clontech Co.) according to 3'RACE. A double-strand cDNA was prepaed from RNA derived from each clone, i. e., poly(A)$^+$RNA of HAS303. 27mer primer OAF075-F1:

5'-CCCCGGGGACATGAGGTGGATACTGTT-3'

containing the deduced initiation ATG codon region based on the information of nucleotide sequence obtained by SST, was prepared. PCR was performed with the said primer and adapter primer attached in the kit. A cDNA which was amplified with clone OAF075B specifically, was recloned by the same method as OC001 to determine full necleotide sequence and obtain

the sequence shown in SEQ ID NO. 14, which was named as OAF075b. An open reading frame was determined and deduced amino acid sequence and nucleotide sequence shown in SEQ ID NOS. 12 and 13, respectively, were obtained.

It was indicated from the results of homology search for the public database of the nucleic acid sequences by using BLASTN and FASTA, and for the public database of the amino acid sequences by using BLASTX, BLASTP and FASTA, that there was no sequence identical to the polypeptide sequence and the nucleotide sequence of OAF075b of the present invention. In addition, the polypeptides of the present invention was expected to possess no transmembrane region by hydrophobisity analysis of the obtained amino acid sequences. From these results, it was proved that polypeptide of the present invention was new secretory protein. Further, the search using BLASTX, BLASTP and FASTA revealed a significant homology between clone OAF075b (region of 1st~359th amino acid in SEQ ID NO. 12) and preprocarboxypeptidase A2 [Homo sapiens] (region of 1st~355th amino acid of Genbank Accession U19977). Based on these homologies, clone OAF075b and preprocarboxypeptidase A2 [Homo sapiens] were expected to share at least some activity.

[Sequence List]

SEQ ID NO. : 1

Length : 344

Type : amino acid

Topology : liner

Molecule type : protein

Sequence Description

```
Met Lys Thr Ile Gln Pro Lys Met His Asn Ser Ile Ser Trp Ala Ile
-28         -25                 -20                 -15
Phe Thr Gly Leu Ala Ala Leu Cys Leu Phe Gln Gly Val Pro Val Arg
        -10                 -5                  1
Ser Gly Asp Ala Thr Phe Pro Lys Ala Met Asp Asn Val Thr Val Arg
  5                 10                  15                  20
Gln Gly Glu Ser Ala Thr Leu Arg Cys Thr Ile Asp Asn Arg Val Thr
                25                  30                  35
Arg Val Ala Trp Leu Asn Arg Ser Thr Ile Leu Tyr Ala Gly Asn Asp
            40                  45                  50
Lys Trp Cys Leu Asp Pro Arg Val Val Leu Leu Ser Asn Thr Gln Thr
        55                  60                  65
Gln Tyr Ser Ile Glu Ile Gln Asn Val Asp Val Tyr Asp Glu Gly Pro
     70                  75                  80
Tyr Thr Cys Ser Val Gln Thr Asp Asn His Pro Lys Thr Ser Arg Val
 85                  90                  95                  100
His Leu Ile Val Gln Val Ser Pro Lys Ile Val Glu Ile Ser Ser Asp
                105                 110                 115
Ile Ser Ile Asn Glu Gly Asn Asn Ile Ser Leu Thr Cys Ile Ala Thr
            120                 125                 130
Gly Arg Pro Glu Pro Thr Val Thr Trp Arg His Ile Ser Pro Lys Ala
```

135 140 145

Val Gly Phe Val Ser Glu Asp Glu Tyr Leu Glu Ile Gln Gly Ile Thr

150 155 160

Arg Glu Gln Ser Gly Asp Tyr Glu Cys Ser Ala Ser Asn Asp Val Ala

165 170 175 180

Ala Pro Val Val Arg Arg Val Lys Val Thr Val Asn Tyr Pro Pro Tyr

185 190 195

Ile Ser Glu Ala Lys Gly Thr Gly Val Pro Val Gly Gln Lys Gly Thr

200 205 210

Leu Gln Cys Glu Ala Ser Ala Val Pro Ser Ala Glu Phe Gln Trp Tyr

215 220 225

Lys Asp Asp Lys Arg Leu Ile Glu Gly Lys Lys Gly Val Lys Val Glu

230 235 240

Asn Arg Pro Phe Leu Ser Lys Leu Ile Phe Phe Asn Val Ser Glu His

245 250 255 260

Asp Tyr Gly Asn Tyr Thr Cys Val Ala Ser Asn Lys Leu Gly His Thr

265 270 275

Asn Ala Ser Ile Met Leu Phe Gly Pro Gly Ala Val Ser Glu Val Ser

280 285 290

Asn Gly Thr Ser Arg Arg Ala Gly Cys Val Trp Leu Leu Pro Leu Leu

295 300 305

Val Leu His Leu Leu Leu Lys Phe

310 315

SEQ ID NO. : 2

Length : 1032

Type : nucleic acid

Strandness : single

Topology : liner

Molecule type : cDNA to mRNA

Sequence Description

ATGAAAACCA TCCAGCCAAA AATGCACAAT TCTATCTCTT GGGCAATCTT CACGGGGCTG 60

GCTGCTCTGT GTCTCTTCCA AGGAGTGCCC GTGCGCAGCG GAGATGCCAC CTTCCCCAAA 120

GCTATGGACA ACGTGACGGT CCGGCAGGGG GAGAGCGCCA CCCTCAGGTG CACTATTGAC 180

AACCGGGTCA CCCGGGTGGC CTGGCTAAAC CGCAGCACCA TCCTCTATGC TGGGAATGAC 240

AAGTGGTGCC TGGATCCTCG CGTGGTCCTT CTGAGCAACA CCCAAACGCA GTACAGCATC 300

GAGATCCAGA ACGTGGATGT GTATGACGAG GGCCCTTACA CCTGCTCGGT GCAGACAGAC 360

AACCACCCAA AGACCTCTAG GGTCCACCTC ATTGTGCAAG TATCTCCCAA AATTGTAGAG 420

ATTTCTTCAG ATATCTCCAT TAATGAAGGG AACAATATTA GCCTCACCTG CATAGCAACT 480

GGTAGACCAG AGCCTACGGT TACTTGGAGA CACATCTCTC CCAAAGCGGT TGGCTTTGTG 540

AGTGAAGACG AATACTTGGA AATTCAGGGC ATCACCCGGG AGCAGTCAGG GGACTACGAG 600

TGCAGTGCCT CCAATGACGT GGCCGCGCCC GTGGTACGGA GAGTAAAGGT CACCGTGAAC 660

TATCCACCAT ACATTTCAGA AGCCAAGGGT ACAGGTGTCC CCGTGGGACA AAAGGGGACA 720

CTGCAGTGTG AAGCCTCAGC AGTCCCCTCA GCAGAATTCC AGTGGTACAA GGATGACAAA 780

AGACTGATTG AAGGAAAGAA AGGGGTGAAA GTGGAAAACA GACCTTTCCT CTCAAAACTC 840

ATCTTCTTCA ATGTCTCTGA ACATGACTAT GGGAACTACA CTTGCGTGGC CTCCAACAAG 900

CTGGGCCACA CCAATGCCAG CATCATGCTA TTTGGTCCAG GCGCCGTCAG CGAGGTGAGC 960

AACGGCACGT CGAGGAGGGC AGGCTGCGTC TGGCTGCTGC CTCTTCTGGT CTTGCACCTG 1020

CTTCTCAAAT TT 1032

SEQ ID NO. : 3

Length : 1693

Type : nucleic acid

Strandness : single

Topology : liner

Molecule type : cDNA to mRNA

Original source

Organism : Homo sapiens

Organelle : human adult brain tissue

Clone Name : OC001

Sequence Description Feature

Name/Key : CDS

Location : 130..1161

Identification method : S

Name/Key : sig peptide

Location : 130..213

Identification method : S

Name/Key : mat peptide

Location : 214..1161

Identification method : S

Sequence Description

```
GTCCTTCAGC AAAACAGTGG ATTTAAATCT CCTTGCACAA GCTTGAGAGC AACACAATCT    60
ATCAGGAAAG AAAGAAAGAA AAAAAACCGA ACCTGACAAA AAAGAAGAAA AAGAAGAAGA   120
AAAAAAATC ATG AAA ACC ATC CAG CCA AAA ATG CAC AAT TCT ATC TCT        168
          Met Lys Thr Ile Gln Pro Lys Met His Asn Ser Ile Ser
          -28         -25                 -20
TGG GCA ATC TTC ACG GGG CTG GCT GCT CTG TGT CTC TTC CAA GGA GTG     216
Trp Ala Ile Phe Thr Gly Leu Ala Ala Leu Cys Leu Phe Gln Gly Val
-15                 -10                 -5                  1
CCC GTG CGC AGC GGA GAT GCC ACC TTC CCC AAA GCT ATG GAC AAC GTG     264
Pro Val Arg Ser Gly Asp Ala Thr Phe Pro Lys Ala Met Asp Asn Val
              5                  10                  15
ACG GTC CGG CAG GGG GAG AGC GCC ACC CTC AGG TGC ACT ATT GAC AAC     312
Thr Val Arg Gln Gly Glu Ser Ala Thr Leu Arg Cys Thr Ile Asp Asn
         20                  25                  30
```

```
CGG GTC ACC CGG GTG GCC TGG CTA AAC CGC AGC ACC ATC CTC TAT GCT     360
Arg Val Thr Arg Val Ala Trp Leu Asn Arg Ser Thr Ile Leu Tyr Ala
     35                  40                  45
GGG AAT GAC AAG TGG TGC CTG GAT CCT CGC GTG GTC CTT CTG AGC AAC     408
Gly Asn Asp Lys Trp Cys Leu Asp Pro Arg Val Val Leu Leu Ser Asn
 50                  55                  60                  65
ACC CAA ACG CAG TAC AGC ATC GAG ATC CAG AAC GTG GAT GTG TAT GAC     456
Thr Gln Thr Gln Tyr Ser Ile Glu Ile Gln Asn Val Asp Val Tyr Asp
                 70                  75                  80
GAG GGC CCT TAC ACC TGC TCG GTG CAG ACA GAC AAC CAC CCA AAG ACC     504
Glu Gly Pro Tyr Thr Cys Ser Val Gln Thr Asp Asn His Pro Lys Thr
             85                  90                  95
TCT AGG GTC CAC CTC ATT GTG CAA GTA TCT CCC AAA ATT GTA GAG ATT     552
Ser Arg Val His Leu Ile Val Gln Val Ser Pro Lys Ile Val Glu Ile
        100                 105                 110
TCT TCA GAT ATC TCC ATT AAT GAA GGG AAC AAT ATT AGC CTC ACC TGC     600
Ser Ser Asp Ile Ser Ile Asn Glu Gly Asn Asn Ile Ser Leu Thr Cys
    115                 120                 125
ATA GCA ACT GGT AGA CCA GAG CCT ACG GTT ACT TGG AGA CAC ATC TCT     648
Ile Ala Thr Gly Arg Pro Glu Pro Thr Val Thr Trp Arg His Ile Ser
130                 135                 140                 145
CCC AAA GCG GTT GGC TTT GTG AGT GAA GAC GAA TAC TTG GAA ATT CAG     696
Pro Lys Ala Val Gly Phe Val Ser Glu Asp Glu Tyr Leu Glu Ile Gln
                150                 155                 160
GGC ATC ACC CGG GAG CAG TCA GGG GAC TAC GAG TGC AGT GCC TCC AAT     744
Gly Ile Thr Arg Glu Gln Ser Gly Asp Tyr Glu Cys Ser Ala Ser Asn
            165                 170                 175
GAC GTG GCC GCG CCC GTG GTA CGG AGA GTA AAG GTC ACC GTG AAC TAT     792
Asp Val Ala Ala Pro Val Val Arg Arg Val Lys Val Thr Val Asn Tyr
```

```
        180                 185                 190
CCA CCA TAC ATT TCA GAA GCC AAG GGT ACA GGT GTC CCC GTG GGA CAA     840
Pro Pro Tyr Ile Ser Glu Ala Lys Gly Thr Gly Val Pro Val Gly Gln
    195                 200                 205
AAG GGG ACA CTG CAG TGT GAA GCC TCA GCA GTC CCC TCA GCA GAA TTC     888
Lys Gly Thr Leu Gln Cys Glu Ala Ser Ala Val Pro Ser Ala Glu Phe
210                 215                 220                 225
CAG TGG TAC AAG GAT GAC AAA AGA CTG ATT GAA GGA AAG AAA GGG GTG     936
Gln Trp Tyr Lys Asp Asp Lys Arg Leu Ile Glu Gly Lys Lys Gly Val
                230                 235                 240
AAA GTG GAA AAC AGA CCT TTC CTC TCA AAA CTC ATC TTC TTC AAT GTC     984
Lys Val Glu Asn Arg Pro Phe Leu Ser Lys Leu Ile Phe Phe Asn Val
            245                 250                 255
TCT GAA CAT GAC TAT GGG AAC TAC ACT TGC GTG GCC TCC AAC AAG CTG    1032
Ser Glu His Asp Tyr Gly Asn Tyr Thr Cys Val Ala Ser Asn Lys Leu
        260                 265                 270
GGC CAC ACC AAT GCC AGC ATC ATG CTA TTT GGT CCA GGC GCC GTC AGC    1080
Gly His Thr Asn Ala Ser Ile Met Leu Phe Gly Pro Gly Ala Val Ser
    275                 280                 285
GAG GTG AGC AAC GGC ACG TCG AGG AGG GCA GGC TGC GTC TGG CTG CTG    1128
Glu Val Ser Asn Gly Thr Ser Arg Arg Ala Gly Cys Val Trp Leu Leu
290                 295                 300                 305
CCT CTT CTG GTC TTG CAC CTG CTT CTC AAA TTT TGATGTGAGT GCCACTTCCC  1181
Pro Leu Leu Val Leu His Leu Leu Leu Lys Phe
                310                 315
CACCCGGGAA AGGCTGCCGC CACCACCACC ACCAACACAA CAGCAATGGC AACACCGACA  1241
GCAACCAATC AGATATATAC AAATGAAATT AGAAGAAACA CAGCCTCATG GGACAGAAAT  1301
TTGAGGGAGG GGAACAAAGA ATACTTTGGG GGGAAAAAAG TTTTAAAAAA GAAATTGAAA  1361
ATTGCCTTGC AGATATTTAG GTACAATGGA GTTTTCTTTT CCCAAACGGG AAGAACACAG  1421
```

CACACCCGGC TTGGACCCAC TGCAAGCTGC ATCGTGCAAC CTCTTTGGTG CCAGTGTGGG 1481

CAAGGGCTCA GCCTCTCTGC CCACAGAGTG CCCCCACGTG GAACATTCTG GAGCTGGCCA 1541

TCCCAAATTC AATCAGTCCA TAGAGACGAA CAGAATGAGA CCTTCCGGCC CAAGCGTGGC 1601

GCTGCGGGCA CTTTGGTAGA CTGTGCCACC ACGGCGTGTG TTGTGAAACG TGAAATAAAA 1661

AGAGCAAAAA AAAAAAAAAA AAAAAAAAAA AA 1693

SEQ ID NO. : 4

Length : 313

Type : amino acid

Topology : liner

Molecule type : protein

Sequence Description

Arg Ser Gly Asp Ala Thr Phe Pro Lys Ala Met Asp Asn Val Thr Val
1 5 10 15

Arg Gln Gly Glu Ser Ala Thr Leu Arg Cys Thr Ile Asp Asn Arg Val
20 25 30

Thr Arg Val Ala Trp Leu Asn Arg Ser Thr Ile Leu Tyr Ala Gly Asn
35 40 45

Asp Lys Trp Cys Leu Asp Pro Arg Val Val Leu Leu Ser Asn Thr Gln
50 55 60

Thr Gln Tyr Ser Ile Glu Ile Gln Asn Val Asp Val Tyr Asp Glu Gly
65 70 75 80

Pro Tyr Thr Cys Ser Val Gln Thr Asp Asn His Pro Lys Thr Ser Arg
85 90 95

Val His Leu Ile Val Gln Val Ser Pro Lys Ile Val Glu Ile Ser Ser
100 105 110

Asp Ile Ser Ile Asn Glu Gly Asn Asn Ile Ser Leu Thr Cys Ile Ala
115 120 125

Thr Gly Arg Pro Glu Pro Thr Val Thr Trp Arg His Ile Ser Pro Lys

130 135 140

Ala Val Gly Phe Val Ser Glu Asp Glu Tyr Leu Glu Ile Gln Gly Ile

145 150 155 160

Thr Arg Glu Gln Ser Gly Asp Tyr Glu Cys Ser Ala Ser Asn Asp Val

165 170 175

Ala Ala Pro Val Val Arg Arg Val Lys Val Thr Val Asn Tyr Pro Pro

180 185 190

Tyr Ile Ser Glu Ala Lys Gly Thr Gly Val Pro Val Gly Gln Lys Gly

195 200 205

Thr Leu Gln Cys Glu Ala Ser Ala Val Pro Ser Ala Glu Phe Gln Trp

210 215 220

Tyr Lys Asp Asp Lys Arg Leu Ile Glu Gly Lys Lys Gly Val Lys Val

225 230 235 240

Glu Asn Arg Pro Phe Leu Ser Lys Leu Ile Phe Phe Asn Val Ser Glu

245 250 255

His Asp Tyr Gly Asn Tyr Thr Cys Val Ala Ser Asn Lys Leu Gly His

260 265 270

Thr Asn Ala Ser Ile Met Leu Phe Gly Pro Gly Ala Val Ser Glu Val

275 280 285

Ser Asn Gly Thr Ser Arg Arg Ala Gly Cys Val Trp Leu Leu Pro Leu

290 295 300

Leu Val Leu His Leu Leu Leu Lys Phe

305 310

SEQ ID NO. : 5

Length : 939

Type : nucleic acid

Strandness : single

Topology : liner

Molecule type : cDNA to mRNA

Sequence Description

CGCAGCGGAG ATGCCACCTT CCCCAAAGCT ATGGACAACG TGACGGTCCG GCAGGGGGAG 60

AGCGCCACCC TCAGGTGCAC TATTGACAAC CGGGTCACCC GGGTGGCCTG GCTAAACCGC 120

AGCACCATCC TCTATGCTGG GAATGACAAG TGGTGCCTGG ATCCTCGCGT GGTCCTTCTG 180

AGCAACACCC AAACGCAGTA CAGCATCGAG ATCCAGAACG TGGATGTGTA TGACGAGGGC 240

CCTTACACCT GCTCGGTGCA GACAGACAAC CACCCAAAGA CCTCTAGGGT CCACCTCATT 300

GTGCAAGTAT CTCCCAAAAT TGTAGAGATT TCTTCAGATA TCTCCATTAA TGAAGGGAAC 360

AATATTAGCC TCACCTGCAT AGCAACTGGT AGACCAGAGC CTACGGTTAC TTGGAGACAC 420

ATCTCTCCCA AAGCGGTTGG CTTTGTGAGT GAAGACGAAT ACTTGGAAAT TCAGGGCATC 480

ACCCGGGAGC AGTCAGGGGA CTACGAGTGC AGTGCCTCCA ATGACGTGGC CGCGCCCGTG 540

GTACGGAGAG TAAAGGTCAC CGTGAACTAT CCACCATACA TTTCAGAAGC CAAGGGTACA 600

GGTGTCCCCG TGGGACAAAA GGGGACACTG CAGTGTGAAG CCTCAGCAGT CCCCTCAGCA 660

GAATTCCAGT GGTACAAGGA TGACAAAAGA CTGATTGAAG GAAAGAAAGG GGTGAAAGTG 720

GAAAACAGAC CTTTCCTCTC AAAACTCATC TTCTTCAATG TCTCTGAACA TGACTATGGG 780

AACTACACTT GCGTGGCCTC CAACAAGCTG GGCCACACCA ATGCCAGCAT CATGCTATTT 840

GGTCCAGGCG CCGTCAGCGA GGTGAGCAAC GGCACGTCGA GGAGGGCAGG CTGCGTCTGG 900

CTGCTGCCTC TTCTGGTCTT GCACCTGCTT CTCAAATTT 939

SEQ ID NO. : 6

Length : 478

Type : amino acid

Strandness : single

Topology : liner

Molecule type : protein

Sequence Description

Met Phe Lys Phe His Gln Met Lys His Ile Phe Glu Ile Leu Asp Lys

1 5 10 15

Met Arg Cys Leu Arg Lys Arg Ser Thr Val Ser Phe Leu Gly Val Leu

20 25 30

Val Ile Phe Leu Leu Phe Met Asn Leu Tyr Ile Glu Asp Ser Tyr Val

35 40 45

Leu Glu Gly Asp Lys Gln Leu Ile Arg Glu Thr Ser Thr His Gln Leu

50 55 60

Asn Ser Glu Arg Tyr Val His Thr Phe Lys Asp Leu Ser Asn Phe Ser

65 70 75 80

Gly Ala Ile Asn Val Thr Tyr Arg Tyr Leu Ala Ala Thr Pro Leu Gln

85 90 95

Arg Lys Arg Tyr Leu Thr Ile Gly Leu Ser Ser Val Lys Arg Lys Lys

100 105 110

Gly Asn Tyr Leu Leu Glu Thr Ile Lys Ser Ile Phe Glu Gln Ser Ser

115 120 125

Tyr Glu Glu Leu Lys Glu Ile Ser Val Val Ile His Leu Ala Asp Phe

130 135 140

Asn Ser Ser Trp Arg Asp Ala Met Val Gln Asp Ile Thr Gln Lys Phe

145 150 155 160

Ala His His Ile Ile Ala Gly Arg Leu Met Val Ile His Ala Pro Glu

165 170 175

Glu Tyr Tyr Pro Ile Leu Asp Gly Leu Lys Arg Asn Tyr Asn Asp Pro

180 185 190

Glu Asp Arg Val Lys Phe Arg Ser Lys Gln Asn Val Asp Tyr Thr Phe

195 200 205

Leu Leu Asn Phe Cys Ala Asn Thr Ser Asp Tyr Tyr Val Met Leu Glu

210 215 220

Asp Asp Val Arg Cys Ser Lys Asn Phe Leu Thr Ala Ile Lys Lys Val

225 230 235 240

Ile Ala Ser Leu Glu Gly Thr Tyr Trp Val Thr Leu Glu Phe Ser Lys

245 250 255

Leu Gly Tyr Ile Gly Lys Leu Tyr His Ser His Asp Leu Pro Arg Leu
260 265 270
Ala His Phe Leu Leu Met Phe Tyr Gln Glu Met Pro Cys Asp Trp Leu
275 280 285
Leu Thr His Phe Arg Gly Leu Leu Ala Gln Lys Asn Val Ile Arg Phe
290 295 300
Lys Pro Ser Leu Phe Gln His Met Gly Tyr Tyr Ser Ser Tyr Lys Gly
305 310 315 320
Thr Glu Asn Lys Leu Lys Asp Asp Asp Phe Glu Glu Glu Ser Phe Asp
325 330 335
Ile Pro Asp Asn Pro Pro Ala Ser Leu Tyr Thr Asn Met Asn Val Phe
340 345 350
Glu Asn Tyr Glu Ala Ser Lys Ala Tyr Ser Ser Val Asp Glu Tyr Phe
355 360 365
Trp Gly Lys Pro Pro Ser Thr Gly Asp Val Phe Val Ile Val Phe Glu
370 375 380
Asn Pro Ile Ile Ile Lys Lys Ile Lys Val Asn Thr Gly Thr Glu Asp
385 390 395 400
Arg Gln Asn Asp Ile Leu His His Gly Ala Leu Asp Val Gly Glu Asn
405 410 415
Val Met Pro Ser Lys Gln Arg Gly Gln Cys Ser Thr Tyr Leu Arg Leu
420 425 430
Gly Glu Phe Lys Asn Gly Asn Phe Glu Met Ser Gly Val Asn Gln Lys
435 440 445
Ile Pro Phe Asp Ile His Cys Met Arg Ile Tyr Val Thr Lys Thr Gln
450 455 460
Lys Glu Trp Leu Ile Ile Arg Ser Ile Ser Ile Trp Thr Ser
465 470 475

SEQ ID NO. : 7

Length : 1434

Type : nucleic acid

Strandness : single

Topology : liner

Molecule type : cDNA to mRNA

Sequence Description

ATGTTTAAAT TTCATCAAAT GAAACATATT TTTGAAATAC TTGATAAAAT GAGATGCCTG 60

AGAAAACGTT CTACAGTGTC ATTCTTGGGA GTTCTTGTCA TTTTTCTCCT TTTTATGAAC 120

TTGTACATTG AAGATAGCTA TGTTCTGGAA GGAGACAAAC AACTTATAAG GGAAACATCC 180

ACACATCAAC TGAATTCAGA ACGCTATGTT CATACTTTCA AGGATTTATC TAATTTCTCA 240

GGAGCCATAA ATGTCACCTA TCGCTACCTA GCTGCCACAC CTTTACAAAG AAAGCGGTAT 300

CTTACAATTG GACTTTCTTC AGTAAAGCGA AAAAAAGGAA ACTATTTACT TGAGACAATT 360

AAGTCAATTT TTGAGCAATC CAGCTATGAA GAGCTGAAGG AAATTTCAGT GGTGATTCAC 420

CTAGCAGACT TTAATTCTTC CTGGCGTGAT GCCATGGTCC AGGATATTAC ACAGAAATTT 480

GCGCACCATA TTATTGCAGG AAGATTAATG GTTATACATG CTCCAGAGGA GTATTACCCA 540

ATCCTAGATG GCCTTAAAAG AAATTACAAT GATCCAGAAG ATAGAGTCAA ATTTCGTTCC 600

AAGCAAAATG TAGATTATAC TTTTCTGCTT AATTTTTGTG CCAATACTTC AGACTATTAT 660

GTAATGCTTG AAGATGATGT TCGATGTTCA AAAAATTTCT TAACTGCCAT CAAGAAAGTC 720

ATTGCATCCC TAGAAGGAAC TTACTGGGTA ACTCTTGAAT TCTCTAAGCT TGGCTACATT 780

GGTAAACTCT ATCATTCTCA TGATCTCCCA CGTTTGGCCC ATTTTTTATT AATGTTTTAT 840

CAAGAAATGC CTTGTGATTG GCTATTGACT CATTTCCGTG GTCTGTTGGC TCAGAAAAAT 900

GTGATCCGTT TTAAACCATC TCTCTTTCAG CACATGGGCT ATTATTCATC ATACAAAGGG 960

ACGGAGAATA AGCTGAAGGA TGATGATTTT GAAGAGGAGT CATTTGACAT TCCTGATAAC 1020

CCCCCTGCAA GTCTGTACAC CAACATGAAT GTGTTTGAAA ATTATGAAGC AAGCAAGGCT 1080

TACAGTAGTG TTGATGAGTA CTTTTGGGGG AAACCACCTT CAACAGGAGA TGTTTTTGTG 1140

ATTGTATTTG AAAATCCAAT TATAATAAAA AAAATTAAAG TAAATACTGG AACAGAAGAT 1200

CGGCAAAATG ATATTTTGCA TCATGGAGCC CTAGATGTTG GGGAAAACGT TATGCCTAGC 1260

AAACAAAGGG GACAATGTTC TACTTACTTA AGACTAGGAG AATTCAAAAA TGGAAACTTT 1320

GAAATGTCAG GTGTAAATCA AAAAATTCCA TTTGATATAC ATTGTATGAG GATATATGTC 1380

ACCAAAACAC AAAAGGAATG GCTAATTATT AGGAGTATTA GCATTTGGAC TTCT 1434

SEQ ID NO. : 8

Length : 2131

Type : nucleic acid

Strandness : single

Topology : liner

Molecule type : cDNA to mRNA

Original source

Organism : Homo sapiens

Organelle : human adult brain tissue

Clone Name : OM237

Sequence Description Feature

Name/Key : CDS

Location : 114..1547

Identification method : S

Sequence Description

CCAGAAAGCA CAGCCCTGAT TCTGCGTGAG AAAGGCTATC TCTACAGAAA CTAAAACGGT 60

ATCAACGGTT TCTGTACAGC ACAGATTATG ACAGCGTCTT TCTTAAGAAG AGA ATG 116

Met

1

TTT AAA TTT CAT CAA ATG AAA CAT ATT TTT GAA ATA CTT GAT AAA ATG 164

Phe Lys Phe His Gln Met Lys His Ile Phe Glu Ile Leu Asp Lys Met

5 10 15

AGA TGC CTG AGA AAA CGT TCT ACA GTG TCA TTC TTG GGA GTT CTT GTC 212

Arg Cys Leu Arg Lys Arg Ser Thr Val Ser Phe Leu Gly Val Leu Val

20 25 30

ATT TTT CTC CTT TTT ATG AAC TTG TAC ATT GAA GAT AGC TAT GTT CTG 260

```
Ile Phe Leu Leu Phe Met Asn Leu Tyr Ile Glu Asp Ser Tyr Val Leu
     35                  40                  45
GAA GGA GAC AAA CAA CTT ATA AGG GAA ACA TCC ACA CAT CAA CTG AAT     308
Glu Gly Asp Lys Gln Leu Ile Arg Glu Thr Ser Thr His Gln Leu Asn
 50                  55                  60                  65
TCA GAA CGC TAT GTT CAT ACT TTC AAG GAT TTA TCT AAT TTC TCA GGA     356
Ser Glu Arg Tyr Val His Thr Phe Lys Asp Leu Ser Asn Phe Ser Gly
                 70                  75                  80
GCC ATA AAT GTC ACC TAT CGC TAC CTA GCT GCC ACA CCT TTA CAA AGA     404
Ala Ile Asn Val Thr Tyr Arg Tyr Leu Ala Ala Thr Pro Leu Gln Arg
             85                  90                  95
AAG CGG TAT CTT ACA ATT GGA CTT TCT TCA GTA AAG CGA AAA AAA GGA     452
Lys Arg Tyr Leu Thr Ile Gly Leu Ser Ser Val Lys Arg Lys Lys Gly
        100                 105                 110
AAC TAT TTA CTT GAG ACA ATT AAG TCA ATT TTT GAG CAA TCC AGC TAT     500
Asn Tyr Leu Leu Glu Thr Ile Lys Ser Ile Phe Glu Gln Ser Ser Tyr
    115                 120                 125
GAA GAG CTG AAG GAA ATT TCA GTG GTG ATT CAC CTA GCA GAC TTT AAT     548
Glu Glu Leu Lys Glu Ile Ser Val Val Ile His Leu Ala Asp Phe Asn
130                 135                 140                 145
TCT TCC TGG CGT GAT GCC ATG GTC CAG GAT ATT ACA CAG AAA TTT GCG     596
Ser Ser Trp Arg Asp Ala Met Val Gln Asp Ile Thr Gln Lys Phe Ala
                150                 155                 160
CAC CAT ATT ATT GCA GGA AGA TTA ATG GTT ATA CAT GCT CCA GAG GAG     644
His His Ile Ile Ala Gly Arg Leu Met Val Ile His Ala Pro Glu Glu
            165                 170                 175
TAT TAC CCA ATC CTA GAT GGC CTT AAA AGA AAT TAC AAT GAT CCA GAA     692
Tyr Tyr Pro Ile Leu Asp Gly Leu Lys Arg Asn Tyr Asn Asp Pro Glu
        180                 185                 190
```

GAT AGA GTC AAA TTT CGT TCC AAG CAA AAT GTA GAT TAT ACT TTT CTG 740
Asp Arg Val Lys Phe Arg Ser Lys Gln Asn Val Asp Tyr Thr Phe Leu
195 200 205

CTT AAT TTT TGT GCC AAT ACT TCA GAC TAT TAT GTA ATG CTT GAA GAT 788
Leu Asn Phe Cys Ala Asn Thr Ser Asp Tyr Tyr Val Met Leu Glu Asp
210 215 220 225

GAT GTT CGA TGT TCA AAA AAT TTC TTA ACT GCC ATC AAG AAA GTC ATT 836
Asp Val Arg Cys Ser Lys Asn Phe Leu Thr Ala Ile Lys Lys Val Ile
230 235 240

GCA TCC CTA GAA GGA ACT TAC TGG GTA ACT CTT GAA TTC TCT AAG CTT 884
Ala Ser Leu Glu Gly Thr Tyr Trp Val Thr Leu Glu Phe Ser Lys Leu
245 250 255

GGC TAC ATT GGT AAA CTC TAT CAT TCT CAT GAT CTC CCA CGT TTG GCC 932
Gly Tyr Ile Gly Lys Leu Tyr His Ser His Asp Leu Pro Arg Leu Ala
260 265 270

CAT TTT TTA TTA ATG TTT TAT CAA GAA ATG CCT TGT GAT TGG CTA TTG 980
His Phe Leu Leu Met Phe Tyr Gln Glu Met Pro Cys Asp Trp Leu Leu
275 280 285

ACT CAT TTC CGT GGT CTG TTG GCT CAG AAA AAT GTG ATC CGT TTT AAA 1028
Thr His Phe Arg Gly Leu Leu Ala Gln Lys Asn Val Ile Arg Phe Lys
290 295 300 305

CCA TCT CTC TTT CAG CAC ATG GGC TAT TAT TCA TCA TAC AAA GGG ACG 1076
Pro Ser Leu Phe Gln His Met Gly Tyr Tyr Ser Ser Tyr Lys Gly Thr
310 315 320

GAG AAT AAG CTG AAG GAT GAT GAT TTT GAA GAG GAG TCA TTT GAC ATT 1124
Glu Asn Lys Leu Lys Asp Asp Asp Phe Glu Glu Glu Ser Phe Asp Ile
325 330 335

CCT GAT AAC CCC CCT GCA AGT CTG TAC ACC AAC ATG AAT GTG TTT GAA 1172
Pro Asp Asn Pro Pro Ala Ser Leu Tyr Thr Asn Met Asn Val Phe Glu

```
        340                 345                 350
AAT TAT GAA GCA AGC AAG GCT TAC AGT AGT GTT GAT GAG TAC TTT TGG    1220
Asn Tyr Glu Ala Ser Lys Ala Tyr Ser Ser Val Asp Glu Tyr Phe Trp
    355                 360                 365
GGG AAA CCA CCT TCA ACA GGA GAT GTT TTT GTG ATT GTA TTT GAA AAT    1268
Gly Lys Pro Pro Ser Thr Gly Asp Val Phe Val Ile Val Phe Glu Asn
370                 375                 380                 385
CCA ATT ATA ATA AAA AAA ATT AAA GTA AAT ACT GGA ACA GAA GAT CGG    1316
Pro Ile Ile Ile Lys Lys Ile Lys Val Asn Thr Gly Thr Glu Asp Arg
                390                 395                 400
CAA AAT GAT ATT TTG CAT CAT GGA GCC CTA GAT GTT GGG GAA AAC GTT    1364
Gln Asn Asp Ile Leu His His Gly Ala Leu Asp Val Gly Glu Asn Val
            405                 410                 415
ATG CCT AGC AAA CAA AGG GGA CAA TGT TCT ACT TAC TTA AGA CTA GGA    1412
Met Pro Ser Lys Gln Arg Gly Gln Cys Ser Thr Tyr Leu Arg Leu Gly
        420                 425                 430
GAA TTC AAA AAT GGA AAC TTT GAA ATG TCA GGT GTA AAT CAA AAA ATT    1460
Glu Phe Lys Asn Gly Asn Phe Glu Met Ser Gly Val Asn Gln Lys Ile
    435                 440                 445
CCA TTT GAT ATA CAT TGT ATG AGG ATA TAT GTC ACC AAA ACA CAA AAG    1508
Pro Phe Asp Ile His Cys Met Arg Ile Tyr Val Thr Lys Thr Gln Lys
450                 455                 460                 465
GAA TGG CTA ATT ATT AGG AGT ATT AGC ATT TGG ACT TCT TAGCCAATTA     1557
Glu Trp Leu Ile Ile Arg Ser Ile Ser Ile Trp Thr Ser
                470                 475
AATCAGTATG TTCAGTTTCT GAAGCAGTTC TTCCTGCTTC GTCTTTTGCT ACCTTTGTCT  1617
TTTGGAGGGA AAGCAATGGA TGGGATATGT TAAAAGAAAC ATTAATTACA TTGGCAGTTT  1677
TCATTTATAC ATTGTTGACA TAATTTTACT CTTAATACAC ACTTGTATTT ATTTTAACGT  1737
CTGAAGTTGA ATATCAGTCT ATAGCTAATG CTACTTTCAT TTATATTTTT AAATGTTCTT  1797
```

```
AGTTTTAAAA TTTCAACTGA TTGTCGAAAG GGTAATATGA AAGATTTTAA ATGAAAAAAA  1857
TTTGTTGGAT GATGATTTTT GAAAAATAGT CACCAACTGT ATATACTTCC TCAAGAACTG  1917
ATAATTCATT ATATCATCAG ATAGCTTTTA TTAAGCATCT GTGGGAATAT ACAGTTGGGT  1977
GGAATGATAA TCTGGTTTAT TTTTTCTGTA AACTTAAGTT TCCGTTGACT TCTGTACATC  2037
TACAATGAAT ACCTCCTCAT AGAAGTGGTG TCTTTACATA ATTTTTTGTG TAGGTGACAC  2097
TATGGAAAAA AAAAAAAAAA AAAAAAAAAA AAAA                               2131
```

SEQ ID NO. : 9

Length : 335

Type : amino acid

Strandness : single

Topology : liner

Molecule type : protein

Sequence Description

```
Met Asp Ser Ala Leu Ser Asp Pro His Asn Gly Ser Ala Glu Ala Gly
 1               5                  10                  15
Gly Pro Thr Asn Ser Thr Thr Arg Pro Pro Ser Thr Pro Glu Gly Ile
            20                  25                  30
Ala Leu Ala Tyr Gly Ser Leu Leu Leu Met Ala Leu Leu Pro Ile Phe
        35                  40                  45
Phe Gly Ala Leu Arg Ser Val Arg Cys Ala Arg Gly Lys Asn Ala Ser
    50                  55                  60
Asp Met Pro Glu Thr Ile Thr Ser Arg Asp Ala Ala Arg Phe Pro Ile
65                  70                  75                  80
Ile Ala Ser Cys Thr Leu Leu Gly Leu Tyr Leu Phe Phe Lys Ile Phe
                85                  90                  95
Ser Gln Glu Tyr Ile Asn Leu Leu Leu Ser Met Tyr Phe Phe Val Leu
            100                 105                 110
Gly Ile Leu Ala Leu Ser His Thr Ile Ser Pro Phe Met Asn Lys Phe
```

115 120 125

Phe Pro Ala Ser Phe Pro Asn Arg Gln Tyr Gln Leu Leu Phe Thr Gln

130 135 140

Gly Ser Gly Glu Asn Lys Glu Glu Ile Ile Asn Tyr Glu Phe Asp Thr

145 150 155 160

Lys Asp Leu Val Cys Leu Gly Leu Ser Ser Ile Val Gly Val Trp Tyr

165 170 175

Leu Leu Arg Lys Val Phe Gly Thr Asn Val Met Val Thr Val Ala Lys

180 185 190

Ser Phe Glu Ala Pro Ile Lys Leu Val Phe Pro Gln Asp Leu Leu Glu

195 200 205

Lys Gly Leu Glu Ala Asn Asn Phe Ala Met Leu Gly Leu Gly Asp Val

210 215 220

Val Ile Pro Gly Ile Phe Ile Ala Leu Leu Leu Arg Phe Asp Ile Ser

225 230 235 240

Leu Lys Lys Asn Thr His Thr Tyr Phe Tyr Thr Ser Phe Ala Ala Tyr

245 250 255

Ile Phe Gly Leu Gly Leu Thr Ile Phe Ile Met His Ile Phe Lys His

260 265 270

Ala Gln Pro Ala Leu Leu Tyr Leu Val Pro Ala Cys Ile Gly Phe Pro

275 280 285

Val Leu Val Ala Leu Ala Lys Gly Glu Val Thr Glu Met Phe Ser Tyr

290 295 300

Glu Glu Ser Asn Pro Lys Asp Pro Ala Ala Val Thr Glu Ser Lys Glu

305 310 315 320

Gly Thr Glu Ala Ser Ala Ser Lys Gly Leu Glu Lys Lys Glu Lys

325 330 335

SEQ ID NO. : 10

Length : 1005

Type : nucleic acid

Strandness : single

Topology : liner

Molecule type : cDNA to mRNA

Sequence Description

ATGGACTCGG CCCTCAGCGA TCCGCATAAC GGCAGTGCCG AGGCAGGCGG CCCCACCAAC 60

AGCACTACGC GGCCGCCTTC CACGCCCGAG GGCATCGCGC TGGCCTACGG CAGCCTCCTG 120

CTCATGGCGC TGCTGCCCAT CTTCTTCGGC GCCCTGCGCT CCGTACGCTG CGCCCGCGGC 180

AAGAATGCTT CAGACATGCC TGAAACAATC ACCAGCCGGG ATGCCGCCCG CTTCCCCATC 240

ATCGCCAGCT GCACACTCTT GGGGCTCTAC CTCTTTTTCA AAATATTCTC CCAGGAGTAC 300

ATCAACCTCC TGCTGTCCAT GTATTTCTTC GTGCTGGGAA TCCTGGCCCT GTCCCACACC 360

ATCAGCCCCT TCATGAATAA GTTTTTTCCA GCCAGCTTTC CAAATCGACA GTACCAGCTG 420

CTCTTCACAC AGGGTTCTGG GGAAAACAAG GAAGAGATCA TCAATTATGA ATTTGACACC 480

AAGGACCTGG TGTGCCTGGG CCTGAGCAGC ATCGTTGGCG TCTGGTACCT GCTGAGGAAG 540

GTATTTGGCA CCAATGTGAT GGTGACAGTG GCCAAGTCCT TCGAGGCACC AATAAAATTG 600

GTGTTTCCCC AGGATCTGCT GGAGAAAGGC CTCGAAGCAA ACAACTTTGC CATGCTGGGA 660

CTTGGAGATG TCGTCATTCC AGGGATCTTC ATTGCCTTGC TGCTGCGCTT TGACATCAGC 720

TTGAAGAAGA ATACCCACAC CTACTTCTAC ACCAGCTTTG CAGCCTACAT CTTCGGCCTG 780

GGCCTTACCA TCTTCATCAT GCACATCTTC AAGCATGCTC AGCCTGCCCT CCTATACCTG 840

GTCCCCGCCT GCATCGGTTT TCCTGTCCTG GTGGCGCTGG CCAAGGGAGA AGTGACAGAG 900

ATGTTCAGTT ATGAGGAGTC AAATCCTAAG GATCCAGCGG CAGTGACAGA ATCCAAAGAG 960

GGAACAGAGG CATCAGCATC GAAGGGGCTG GAGAAGAAAG AGAAA 1005

SEQ ID NO. : 11

Length : 1486

Type : nucleic acid

Strandness : single

Topology : liner

Molecule type : cDNA to mRNA

Original source

Organism : Homo sapiens

Cell line : T98G

Clone Name : OA004b

Sequence Description Feature

Name/Key : CDS

Location : 117..1121

Identification method : S

Sequence Description

```
CACGTCACTT CCTGTTGCCT TAGGGGAACG TGGCTTTCCC TGCAGAGCCG GTGTCTCCGC    60
CTGCGTCCCT GCTGCAGCAA CCGGAGCTGG AGTCGGATCC CGAACGCACC CTCGCC       116
ATG GAC TCG GCC CTC AGC GAT CCG CAT AAC GGC AGT GCC GAG GCA GGC      164
Met Asp Ser Ala Leu Ser Asp Pro His Asn Gly Ser Ala Glu Ala Gly
  1               5                  10                  15
GGC CCC ACC AAC AGC ACT ACG CGG CCG CCT TCC ACG CCC GAG GGC ATC      212
Gly Pro Thr Asn Ser Thr Thr Arg Pro Pro Ser Thr Pro Glu Gly Ile
             20                  25                  30
GCG CTG GCC TAC GGC AGC CTC CTG CTC ATG GCG CTG CTG CCC ATC TTC      260
Ala Leu Ala Tyr Gly Ser Leu Leu Leu Met Ala Leu Leu Pro Ile Phe
         35                  40                  45
TTC GGC GCC CTG CGC TCC GTA CGC TGC GCC CGC GGC AAG AAT GCT TCA      308
Phe Gly Ala Leu Arg Ser Val Arg Cys Ala Arg Gly Lys Asn Ala Ser
     50                  55                  60
GAC ATG CCT GAA ACA ATC ACC AGC CGG GAT GCC GCC CGC TTC CCC ATC      356
Asp Met Pro Glu Thr Ile Thr Ser Arg Asp Ala Ala Arg Phe Pro Ile
 65                  70                  75                  80
ATC GCC AGC TGC ACA CTC TTG GGG CTC TAC CTC TTT TTC AAA ATA TTC      404
Ile Ala Ser Cys Thr Leu Leu Gly Leu Tyr Leu Phe Phe Lys Ile Phe
```

```
                    85                  90                  95
TCC CAG GAG TAC ATC AAC CTC CTG CTG TCC ATG TAT TTC TTC GTG CTG     452
Ser Gln Glu Tyr Ile Asn Leu Leu Leu Ser Met Tyr Phe Phe Val Leu
                100                 105                 110
GGA ATC CTG GCC CTG TCC CAC ACC ATC AGC CCC TTC ATG AAT AAG TTT     500
Gly Ile Leu Ala Leu Ser His Thr Ile Ser Pro Phe Met Asn Lys Phe
        115                 120                 125
TTT CCA GCC AGC TTT CCA AAT CGA CAG TAC CAG CTG CTC TTC ACA CAG     548
Phe Pro Ala Ser Phe Pro Asn Arg Gln Tyr Gln Leu Leu Phe Thr Gln
    130                 135                 140
GGT TCT GGG GAA AAC AAG GAA GAG ATC ATC AAT TAT GAA TTT GAC ACC     596
Gly Ser Gly Glu Asn Lys Glu Glu Ile Ile Asn Tyr Glu Phe Asp Thr
145                 150                 155                 160
AAG GAC CTG GTG TGC CTG GGC CTG AGC AGC ATC GTT GGC GTC TGG TAC     644
Lys Asp Leu Val Cys Leu Gly Leu Ser Ser Ile Val Gly Val Trp Tyr
                165                 170                 175
CTG CTG AGG AAG GTA TTT GGC ACC AAT GTG ATG GTG ACA GTG GCC AAG     692
Leu Leu Arg Lys Val Phe Gly Thr Asn Val Met Val Thr Val Ala Lys
            180                 185                 190
TCC TTC GAG GCA CCA ATA AAA TTG GTG TTT CCC CAG GAT CTG CTG GAG     740
Ser Phe Glu Ala Pro Ile Lys Leu Val Phe Pro Gln Asp Leu Leu Glu
        195                 200                 205
AAA GGC CTC GAA GCA AAC AAC TTT GCC ATG CTG GGA CTT GGA GAT GTC     788
Lys Gly Leu Glu Ala Asn Asn Phe Ala Met Leu Gly Leu Gly Asp Val
    210                 215                 220
GTC ATT CCA GGG ATC TTC ATT GCC TTG CTG CTG CGC TTT GAC ATC AGC     836
Val Ile Pro Gly Ile Phe Ile Ala Leu Leu Leu Arg Phe Asp Ile Ser
225                 230                 235                 240
TTG AAG AAG AAT ACC CAC ACC TAC TTC TAC ACC AGC TTT GCA GCC TAC     884
```

```
Leu Lys Lys Asn Thr His Thr Tyr Phe Tyr Thr Ser Phe Ala Ala Tyr
                245                 250                 255
ATC TTC GGC CTG GGC CTT ACC ATC TTC ATC ATG CAC ATC TTC AAG CAT      932
Ile Phe Gly Leu Gly Leu Thr Ile Phe Ile Met His Ile Phe Lys His
            260                 265                 270
GCT CAG CCT GCC CTC CTA TAC CTG GTC CCC GCC TGC ATC GGT TTT CCT      980
Ala Gln Pro Ala Leu Leu Tyr Leu Val Pro Ala Cys Ile Gly Phe Pro
        275                 280                 285
GTC CTG GTG GCG CTG GCC AAG GGA GAA GTG ACA GAG ATG TTC AGT TAT     1028
Val Leu Val Ala Leu Ala Lys Gly Glu Val Thr Glu Met Phe Ser Tyr
    290                 295                 300
GAG GAG TCA AAT CCT AAG GAT CCA GCG GCA GTG ACA GAA TCC AAA GAG     1076
Glu Glu Ser Asn Pro Lys Asp Pro Ala Ala Val Thr Glu Ser Lys Glu
305                 310                 315                 320
GGA ACA GAG GCA TCA GCA TCG AAG GGG CTG GAG AAG AAA GAG AAA         1121
Gly Thr Glu Ala Ser Ala Ser Lys Gly Leu Glu Lys Lys Glu Lys
                325                 330                 335
TGATGCGGCT GGTGCCCGAG CCTCTCAGGG CCAGACCAGA CAGATGGGGG CTGGGCCCAC   1181
ACAGGCGTGC ACCGGTAGAG GGCACAGGAG GCCAAGGGCA GCTCCAGGAC AGGGCAGGGG   1241
GCAGCAGGAT ACCTCCAGCC AGGCCTCTGT GGCCTCTGTT TCCTTCTCCC TTTCTTGGCC   1301
CTCCTCTGCT CCTCCCCACA CCCTGCAGGC AAAAGAAACC CCCAGCTTCC CCCCTCCCCG   1361
GGAGCCAGGT GGGAAAAGTG GGTGTGATTT TTAGATTTTG TATTGTGGAC TGATTTTGCC   1421
TCACATTAAA AACTCATCCC ATGGCCAGGG CGGGCCACTG TGCTCCTGAA AAAAAAAAAA   1481
AAAAA                                                               1486
```

SEQ ID NO. : 12

Length : 360

Type : amino acid

Strandness : single

Topology : liner

Molecule type : protein

Sequence Description

```
Met Arg Trp Ile Leu Phe Ile Gly Ala Leu Ile Gly Ser Ser Ile Cys
-16 -15                 -10                 -5
Gly Gln Glu Lys Phe Phe Gly Asp Gln Val Phe Arg Ile Asn Val Arg
 1               5                  10                  15
Asn Gly Asp Glu Ile Ser Lys Leu Ser Gln Leu Val Asn Ser Asn Asn
            20                  25                  30
Leu Lys Leu Asn Phe Trp Lys Ser Pro Ser Ser Phe Asn Arg Pro Val
        35                  40                  45
Asp Val Leu Val Pro Ser Val Ser Leu Gln Ala Phe Lys Ser Phe Leu
    50                  55                  60
Arg Ser Gln Gly Leu Glu Tyr Ala Val Thr Ile Glu Asp Leu Gln Ala
65                  70                  75                  80
Leu Leu Asp Asn Glu Asp Asp Glu Met Gln His Asn Glu Gly Gln Glu
                85                  90                  95
Arg Ser Ser Asn Asn Phe Asn Tyr Gly Ala Tyr His Ser Leu Glu Ala
            100                 105                 110
Ile Tyr His Glu Met Asp Asn Ile Ala Ala Asp Phe Pro Asp Leu Ala
        115                 120                 125
Arg Arg Val Lys Ile Gly His Ser Phe Glu Asn Arg Pro Met Tyr Val
    130                 135                 140
Leu Lys Phe Ser Thr Gly Lys Gly Val Arg Arg Pro Ala Val Trp Leu
145                 150                 155                 160
Asn Ala Gly Ile His Ser Arg Glu Trp Ile Ser Gln Ala Thr Ala Ile
                165                 170                 175
Trp Thr Ala Arg Lys Ile Val Ser Asp Tyr Gln Arg Asp Pro Ala Ile
            180                 185                 190
```

Thr Ser Ile Leu Glu Lys Met Asp Ile Phe Leu Leu Pro Val Ala Asn
195 200 205

Pro Asp Gly Tyr Val Tyr Thr Gln Thr Gln Asn Arg Leu Trp Arg Lys
210 215 220

Thr Arg Ser Arg Asn Pro Gly Ser Ser Cys Ile Gly Ala Asp Pro Asn
225 230 235 240

Arg Ser Trp Asn Ala Ser Phe Ala Gly Lys Gly Ala Ser Asp Asn Pro
245 250 255

Cys Ser Glu Val Tyr His Gly Pro His Ala Asn Ser Glu Val Glu Val
260 265 270

Lys Ser Val Val Asp Phe Ile Gln Lys His Gly Asn Phe Lys Cys Phe
275 280 285

Ile Asp Leu His Ser Tyr Ser Gln Leu Leu Met Tyr Pro Tyr Gly Tyr
290 295 300

Ser Val Lys Lys Ala Pro Asp Ala Glu Glu Leu Asp Lys Val Ala Arg
305 310 315 320

Leu Ala Ala Lys Ala Leu Ala Ser Val Ser Gly Thr Glu Tyr Gln Val
325 330 335

Gly Pro Thr Cys Thr Thr Val Leu
340

SEQ ID NO. : 13

Length : 1080

Type : nucleic acid

Strandness : single

Topology : liner

Molecule type : cDNA to mRNA

Sequence Description

ATGAGGTGGA TACTGTTCAT TGGGGCCCTT ATTGGGTCCA GCATCTGTGG CCAAGAAAAA 60

TTTTTTGGGG ACCAAGTTTT TAGGATTAAT GTCAGAAATG GAGACGAGAT CAGCAAATTG 120
AGTCAACTAG TGAATTCAAA CAACTTGAAG CTCAATTTCT GGAAATCTCC CTCCTCCTTC 180
AATCGGCCTG TGGATGTCCT GGTCCCATCT GTCAGTCTGC AGGCATTTAA ATCCTTCCTG 240
AGATCCCAGG GCTTAGAGTA CGCAGTGACA ATTGAGGACC TGCAGGCCCT TTTAGACAAT 300
GAAGATGATG AAATGCAACA CAATGAAGGG CAAGAACGGA GCAGTAATAA CTTCAACTAC 360
GGGGCTTACC ATTCCCTGGA AGCTATTTAC CACGAGATGG ACAACATTGC CGCAGACTTT 420
CCTGACCTGG CGAGGAGGGT GAAGATTGGA CATTCGTTTG AAAACCGGCC GATGTATGTA 480
CTGAAGTTCA GCACTGGGAA AGGCGTGAGG CGGCCGGCCG TTTGGCTGAA TGCAGGCATC 540
CATTCCCGAG AGTGGATCTC CCAGGCCACT GCAATCTGGA CGGCAAGGAA GATTGTATCT 600
GATTACCAGA GGGATCCAGC TATCACCTCC ATCTTGGAGA AAATGGATAT TTTCTTGTTG 660
CCTGTGGCCA ATCCTGATGG ATATGTGTAT ACTCAAACTC AAAACCGATT ATGGAGGAAG 720
ACGCGGTCCC GAAATCCTGG AAGCTCCTGC ATTGGTGCTG ACCCAAATAG AAGCTGGAAC 780
GCTAGTTTTG CAGGAAAGGG AGCCAGCGAC AACCCTTGCT CCGAAGTGTA CCATGGACCC 840
CACGCCAATT CGGAAGTGGA GGTGAAATCA GTGGTAGATT TCATCCAAAA ACATGGGAAT 900
TTCAAGTGCT TCATCGACCT GCACAGCTAC TCGCAGCTGC TGATGTATCC ATATGGGTAC 960
TCAGTCAAAA AGGCCCCAGA TGCCGAGGAA CTCGACAAGG TGGCGAGGCT TGCGGCCAAA 1020
GCTCTGGCTT CTGTGTCGGG CACTGAGTAC CAAGTGGGTC CCACCTGCAC CACTGTCTTA 1080

SEQ ID NO. : 14

Length : 3156

Type : nucleic acid

Strandness : single

Topology : liner

Molecule type : cDNA to mRNA

Original source

Organism : Homo sapiens

Cell line : human bone marrow stroma cell HAS303

Clone Name : OAF075b

Sequence Description Feature

Name/Key : CDS

Location : 11..1090

Identification method : S

Name/Key : sig peptide

Location : 11..58

Identification method : S

Name/Key : mat peptide

Location : 59..1090

Identification method : S

Sequence Description

```
CCCCGGGGAC ATG AGG TGG ATA CTG TTC ATT GGG GCC CTT ATT GGG TCC      49
           Met Arg Trp Ile Leu Phe Ile Gly Ala Leu Ile Gly Ser
           -16 -15                 -10                 -5

AGC ATC TGT GGC CAA GAA AAA TTT TTT GGG GAC CAA GTT TTT AGG ATT      97
Ser Ile Cys Gly Gln Glu Lys Phe Phe Gly Asp Gln Val Phe Arg Ile
               1               5                  10

AAT GTC AGA AAT GGA GAC GAG ATC AGC AAA TTG AGT CAA CTA GTG AAT     145
Asn Val Arg Asn Gly Asp Glu Ile Ser Lys Leu Ser Gln Leu Val Asn
     15                  20                  25

TCA AAC AAC TTG AAG CTC AAT TTC TGG AAA TCT CCC TCC TCC TTC AAT     193
Ser Asn Asn Leu Lys Leu Asn Phe Trp Lys Ser Pro Ser Ser Phe Asn
 30                  35                  40                  45

CGG CCT GTG GAT GTC CTG GTC CCA TCT GTC AGT CTG CAG GCA TTT AAA     241
Arg Pro Val Asp Val Leu Val Pro Ser Val Ser Leu Gln Ala Phe Lys
                50                  55                  60

TCC TTC CTG AGA TCC CAG GGC TTA GAG TAC GCA GTG ACA ATT GAG GAC     289
Ser Phe Leu Arg Ser Gln Gly Leu Glu Tyr Ala Val Thr Ile Glu Asp
```

```
                65                  70                  75
CTG CAG GCC CTT TTA GAC AAT GAA GAT GAT GAA ATG CAA CAC AAT GAA     337
Leu Gln Ala Leu Leu Asp Asn Glu Asp Asp Glu Met Gln His Asn Glu
            80                  85                  90
GGG CAA GAA CGG AGC AGT AAT AAC TTC AAC TAC GGG GCT TAC CAT TCC     385
Gly Gln Glu Arg Ser Ser Asn Asn Phe Asn Tyr Gly Ala Tyr His Ser
        95                  100                 105
CTG GAA GCT ATT TAC CAC GAG ATG GAC AAC ATT GCC GCA GAC TTT CCT     433
Leu Glu Ala Ile Tyr His Glu Met Asp Asn Ile Ala Ala Asp Phe Pro
    110                 115                 120                 125
GAC CTG GCG AGG AGG GTG AAG ATT GGA CAT TCG TTT GAA AAC CGG CCG     481
Asp Leu Ala Arg Arg Val Lys Ile Gly His Ser Phe Glu Asn Arg Pro
                130                 135                 140
ATG TAT GTA CTG AAG TTC AGC ACT GGG AAA GGC GTG AGG CGG CCG GCC     529
Met Tyr Val Leu Lys Phe Ser Thr Gly Lys Gly Val Arg Arg Pro Ala
            145                 150                 155
GTT TGG CTG AAT GCA GGC ATC CAT TCC CGA GAG TGG ATC TCC CAG GCC     577
Val Trp Leu Asn Ala Gly Ile His Ser Arg Glu Trp Ile Ser Gln Ala
        160                 165                 170
ACT GCA ATC TGG ACG GCA AGG AAG ATT GTA TCT GAT TAC CAG AGG GAT     625
Thr Ala Ile Trp Thr Ala Arg Lys Ile Val Ser Asp Tyr Gln Arg Asp
    175                 180                 185
CCA GCT ATC ACC TCC ATC TTG GAG AAA ATG GAT ATT TTC TTG TTG CCT     673
Pro Ala Ile Thr Ser Ile Leu Glu Lys Met Asp Ile Phe Leu Leu Pro
190                 195                 200                 205
GTG GCC AAT CCT GAT GGA TAT GTG TAT ACT CAA ACT CAA AAC CGA TTA     721
Val Ala Asn Pro Asp Gly Tyr Val Tyr Thr Gln Thr Gln Asn Arg Leu
                210                 215                 220
TGG AGG AAG ACG CGG TCC CGA AAT CCT GGA AGC TCC TGC ATT GGT GCT     769
```

```
Trp Arg Lys Thr Arg Ser Arg Asn Pro Gly Ser Ser Cys Ile Gly Ala
            225                 230                 235
GAC CCA AAT AGA AGC TGG AAC GCT AGT TTT GCA GGA AAG GGA GCC AGC     817
Asp Pro Asn Arg Ser Trp Asn Ala Ser Phe Ala Gly Lys Gly Ala Ser
        240                 245                 250
GAC AAC CCT TGC TCC GAA GTG TAC CAT GGA CCC CAC GCC AAT TCG GAA     865
Asp Asn Pro Cys Ser Glu Val Tyr His Gly Pro His Ala Asn Ser Glu
    255                 260                 265
GTG GAG GTG AAA TCA GTG GTA GAT TTC ATC CAA AAA CAT GGG AAT TTC     913
Val Glu Val Lys Ser Val Val Asp Phe Ile Gln Lys His Gly Asn Phe
270                 275                 280                 285
AAG TGC TTC ATC GAC CTG CAC AGC TAC TCG CAG CTG CTG ATG TAT CCA     961
Lys Cys Phe Ile Asp Leu His Ser Tyr Ser Gln Leu Leu Met Tyr Pro
                290                 295                 300
TAT GGG TAC TCA GTC AAA AAG GCC CCA GAT GCC GAG GAA CTC GAC AAG    1009
Tyr Gly Tyr Ser Val Lys Lys Ala Pro Asp Ala Glu Glu Leu Asp Lys
            305                 310                 315
GTG GCG AGG CTT GCG GCC AAA GCT CTG GCT TCT GTG TCG GGC ACT GAG    1057
Val Ala Arg Leu Ala Ala Lys Ala Leu Ala Ser Val Ser Gly Thr Glu
        320                 325                 330
TAC CAA GTG GGT CCC ACC TGC ACC ACT GTC TTA TAAACTGCCA AAACTGGGAG   1110
Tyr Gln Val Gly Pro Thr Cys Thr Thr Val Leu
    335                 340
ATACTCATCA GATTGCTCCA ACAGAAGAGG AGGAAGGCTC TCCCGAGGGC TGTCCAGGAG   1170
ACATAAAAAT TCTACCTTTT CTTTTCTTTT TGAAATGGAG TTTCGTTTCG CTCTTGTTGC   1230
CCAGGCTGGA GTGCAATGGC GTGATCTCCA CTCATCGCAA CTTCCGCCTC CCAGGTTCAA   1290
GCGATTCCCC TGCCTCAGCC TCCCGAGTAA CTGGGATTAT AGGCATGTGC CCCACCCCCA   1350
ACTAATTTTT GTATTTTTAG TAGAGATGGG GTTTCTCCAT GTTGGTCAGT CTGGTCTTGA   1410
GCTCCCGACC TCAGGTGATC TGCCCGCCTC GGCCTCTCAA AGTGCTGGGA TTACAGGCGT   1470
```

GAGCCACAGC ACCCGGCCAA AATGTCCACC TTTTCTAAGA GCCCATCTTC CATATTCTTT 1530

ATAGGCCTTG TCTGTCCTTG TTTTTTCAAA AAAAAAACAA TCAATTTTTG TATAATAGCA 1590

CTCTATCCAA CGCCATAGGT TATGGTGTGT GCTACATACA CAGTCGACGT TTGTCCTTTC 1650

AAGTGCTGGG CCTTTTCCTA GATCGCCATT TTAGAGGAAA ATAATTCTAA AATGGATTTT 1710

ACACTCTTCT GCCTTCTAAA ACAGAGCATG GAGAAGAGAT TTAAGCCCCT TTTTTCATGG 1770

TTAAGTGTAC TTCTCAACCT CAGTTCGTAT ATGCTAAAGG CCTACTCTGC CGTCTTGGAC 1830

TGTTTGGACC TTCTGCTAAA TGATCCTGGC CTGTTTTCCT TCTTGTGTTT GCTTTGTAGA 1890

GTTTTGTGTC TCCTTTCTCC TGCCAGACTG TCAGCAGTAG CTTGTATTGC TTCAGGCCAA 1950

CAGCCTCTAG CAACCCTTTC CCCTCCTCTT CACTGATTCT GCTCCAGGAA GGGCTTGGAA 2010

ACAAGTTCTT TGGGTTCATC TGACTTGTGG ATAACACAGT TTCATGTACT TTTTGTAGTT 2070

CATAAGCGTG GTGATTGGGT TTTCACGCTC ATGTGTGACA TATGCCTTCC TCCAATTTTG 2130

TTACAATGTT GGTGCGTTAC CCATCAGACA TGGGGGAAGA AAGGGTGTTC ATGACAGCAT 2190

TATCCATAGT TACAAAAGAC ATGTACAGGG GCCAAGGGAA AACTTCCCCT TTGCCTTCTG 2250

AAGGTTCATT GAAAAATCAA CTGACCAAAG GCAGATCGAT AGGAGAAAAG GCATACAAAA 2310

TTTTATTTTA GTGTGCATGG CACAGGGGAA TCACAGGAGA ATGATTTCCC AATAACCCAA 2370

TGGGGCACAG AAGCTTGTAT ACCCTTTTTC ATACAGGAGG GAGGAGATGT ATGGACTGGG 2430

GAGGTGGGAG GCAGATATTA CAGGAAGGTG AGGGGCGGAG CTGTACAGGA ACAAAGCTTG 2490

TCTTATTAAG CAGATAAAGT CCTCCAGGCA ATCTCTTGGA GCTGCTCTCA GAAGAATAGA 2550

TGAAGTCTGT CTGGGTGTGG TGATGATTCC CAGTCTCATC TCTTCTGGTG GTTTATCTTT 2610

CTTGGTTATT TGATGAGACC TCTAGGGAGG GTGTTTAAGA CAATTGCATT TCTTTTGGAA 2670

AGATGCTTTC TTGGTCAGAT GAGGAAATTT CCAAAGACAG ACAGTCCCTC CCTGTGTTTG 2730

GTGGTGGGGC AGGTATGGGG AACAAGAAGT TAGAGGGACC TTGGTTCGGG GGCGGCTTCT 2790

GAGGGCCCTC AGCATGTCAA AACATCAGCC TTTGGGATAT CACTTTCTGA GCCCCAACCC 2850

TTGTAAGTGT CTAAAATGTC CACCTAGAGA ATGCAGGATA AATACACATT TGGTGCATTC 2910

ACACAATGCA GCACTACGGA GCCCTTAAAT GAATGAGGTA GATCTATGTG CGCTAAAAGG 2970

GAATACTCAC CAATTGTTAA TTGAAAAATA CATGTGCAGA ACAGCGTTAA TAGTGTGTTC 3030

CCATTTTTTG TTGTTGTTAT TGTTTTTAAA GAGTAGGTAG ACTTTCAGCA GGGACCCAAA 3090

TAAAGTGAAG TTTACAAACT TCGTCATTTT GACTGAAAAA AAAAAAAAAA AAAAAAAAAA 3150

AAAAAA 3156